

GEOGRAPHIC INFORMATION SYSTEMS, SPATIAL DATA  
ANALYSIS, AND DECISION MAKING IN GOVERNMENT

R. F. Tomlinson

July, 1974

Roger Tomlinson completed this PhD thesis at UCL in 1974. It was based on his experience during the 1960s in creating the Canada Geographic Information System (CGIS) for the Canada Land Inventory. The CGIS is now widely recognised as the world's first integrated geographical information system capable of mapping and analysing geographical information and applying it to a comprehensive range of planning and management functions.

Roger had come to UCL as a result of a chance meeting with Professor Terry Coppock at the 1964 International Geographical Union meeting in London. Terry was interested in automating land use mapping, and saw the potential value of documenting the CGIS experience.

By the time contact with UCL was re-established in 1970, Terry had moved on to Edinburgh, and the task of focusing Roger's experience and enthusiasm into an academic thesis fell to me, a young economic geographer with an interest in planning and spatial analysis. The thesis documented Roger's experience in developing the CGIS, laying out the principles required to apply geographical information systems more generally. It was, however, more than an account of technical requirements. Roger always insisted that GIS applications require geographical expertise and should focus on applying geographical knowledge and understanding to the analysis of real world problems.

GIS developments at the time were limited by computing capacities, which now seem extraordinarily limited, but the thesis was able to develop general principles from the land use mapping problems encountered in the CGIS. Roger knew that computing capacity would inevitably expand, although even he, I suspect, did not anticipate the speed with which this would occur after the 1970s. Remarkably, this was when the approaches he set down in the early 1970s came into their own, developed through his seminal book, *Thinking about GIS: Geographical Information System Planning for Managers*, now in its fifth edition and including a translation into Chinese.

From the work summarised in this thesis, Roger's later achievements in the application of GIS were summarised in the May 2001 citation for his award of the Order of Canada:

*Recognized as the 'father of GIS', he pioneered its uses worldwide to collect, manage and manipulated geographical data, changing the face of geography as a discipline. His work established Canada's reputation for excellence in the emerging and continually expanding field of geo-spatial analysis. Governments and scientists around the world have turned to him to better understand our environment and changing patterns of land use, to better manage urban development and our precious natural resources.*

Peter Wood  
July 2017

The application of electronic computing methods  
and techniques to the storage, compilation, and  
assessment of mapped data

Roger F. Tomlinson, B.Sc., M.Sc.

A thesis submitted in candidature  
for the Degree of Doctor of Philosophy

University of London

July, 1974

## ABSTRACT

Manual and computer-aided techniques for handling location-specific data are viewed, not in terms of their ability to produce maps, but in the broader role of storing, manipulating, and displaying such data.

The characteristics, sources, and supply of spatial data are discussed. Manual techniques for handling mapped data are examined as the benchmarks of traditional capability against which the new computer-aided methods must be judged.

The storage of spatial data in electronic computers is reviewed. General categories of computer-aided techniques based on their storage format are examined and their characteristics are illustrated with specific examples. The Canada Geographic Information System, whose development was initiated and directed by the author, is examined in this context.

The current overall ability to handle spatial data is analysed within a framework of data characteristics and system capabilities. The categories of different capabilities emerge, their institutional backgrounds are made clear, but questions are raised about the way in which the techniques are developing, their usefulness, and the future progress of such geographic information systems.

Formal spatial analysis and government decision making are examined in terms of their use of data. The contribution of existing geographic information systems to these processes is critically appraised.

Within the discipline of geography, it is suggested that the mutual development of formal spatial models and geographic information systems will lead to future beneficial shifts of emphasis in both fields of endeavour. Within government, changes in operational procedure and the possibility of changes in the decision-making process are discussed. Organizational structures to facilitate the use of data stores are postulated. It is apparent that the development of geographic information systems cannot sensibly proceed further in isolation, but must be undertaken as an integral part of the very large structure of data gathering, data analysis, and decision making.



## ACKNOWLEDGEMENTS

I would like to thank Dr. P. Wood for the advice and encouragement he so generously provided during the preparation of this thesis.

Appreciation is also extended to Professor W. Mead for the many helpful discussions during my residence at the Department of Geography, University College, London.

The financial assistance provided by the Canada Council is gratefully acknowledged.

To the many colleagues and friends who have individually and jointly taken part in discussions and provided assistance in many ways, I extend my heartfelt thanks.

R. F. T.

Ottawa, 1974

## TABLE OF CONTENTS

	Page
Abstract .....	2
Acknowledgements .....	3
Table of Contents .....	4
List of Figures .....	8
List of Tables .....	11
Chapter 1 <u>Introduction</u> .....	12
Review of literature .....	14
Contribution and context of thesis .....	17
Thesis structure .....	18
Chapter 2 <u>Characteristics and Availability of</u> <u>Location-Specific Data</u> .....	26
Characteristics of location-specific data .....	26
Concepts of space .....	27
Spatial languages .....	28
Spatial description .....	29
Spatial data format .....	30
Types of location identifier .....	34
Supply and availability of data .....	37
Source characteristics .....	37
Perceived need for data .....	37
Gathering constraints .....	40
Gathering capability .....	43
(a) Gathering techniques .....	43
(b) Available resources .....	48
Resulting data characteristics .....	49
Validity constraints .....	50
Physical handling characteristics .....	53
Use constraints .....	55
Resulting spatial data supply and availability .....	56
Summary .....	58
Chapter 3 <u>Manual Techniques for Handling Mapped and Other</u> <u>Location-Specific Data</u> .....	59
Image manipulation .....	61
Scale change .....	62
Distortion elimination .....	62
Nonlinear distortion elimination and map projection change .....	63

	Page
Grouping .....	63
Merge and dissolve .....	64
Edge match .....	64
Centroid allocation .....	64
Contouring .....	65
Generalization .....	65
Data extraction .....	66
Data selection .....	67
Reading .....	67
Search and identification .....	68
Data retrieval .....	69
Transfer .....	69
Measurement and comparison .....	73
Summary .....	78
Chapter 4 <u>Current Computer-Aided Systems for Handling Mapped</u> <u>and Other Location-Specific Data</u> .....	82
Basic considerations .....	84
Codes for non-graphic spatial data .....	85
File structures for coded non-graphic spatial data .....	90
File sequencing .....	91
File linkage .....	92
File record layout .....	94
File information content .....	95
Relationship between type of storage format for non- graphic spatial data and current systems for handling spatial data .....	95
Arbitrary regular-area category of systems .....	98
Arbitrary regular-area systems - Type 1 .....	99
LUNR system - Example of Type 1 of the arbitrary regular-area category .....	99
Arbitrary regular-area systems - Type 2 .....	103
SYMAP system - Example of Type 2 of the arbitrary regular-area category .....	105
Point data category of systems .....	109
GRDSR system - Example of point data category .....	111
Line segment category of systems .....	115
DIMF system - Example of line segment category .....	116
AUTOMAP system - Example of line segment category .....	122
Summary .....	126
Chapter 5 <u>Computer-Aided Systems Under Development for</u> <u>Handling Mapped and Other Location-Specific Data</u> .....	130
Irregular polygon category of systems .....	130

	Page
Types of irregular polygon systems.....	143
Coarse polygon systems: the example of PIOS .....	144
Fine polygon systems: the example of CGIS.....	152
Current hardware-related developments in input, manipulation, and output capabilities .....	168
Input techniques .....	168
Current point/line digitizing .....	168
Interactive point/line digitizing .....	171
Automatic line following.....	172
Scanning.....	173
Data base manipulation techniques.....	175
Interactive manipulation of spatial data base.....	175
Manipulation of ERTS numerically coded images .....	178
Output devices.....	179
Multi-format systems.....	180
Summary .....	184
 Chapter 6 <u>Analysis of Spatial Data Handling Capability in Terms of</u> <u>Data Character and Availability: A Framework for</u> <u>Description of Geographic Information Systems</u> .....	187
Analysis of framework .....	187
Location identifiers.....	188
Volume of related data .....	192
Manipulation facilities .....	194
Synthesis .....	210
 Chapter 7 <u>Geographic Information Systems, Spatial Data Analysis,</u> <u>and Decision Making in Government</u> .....	221
Formal spatial analysis related to data use.....	223
Government decision making related to data use.....	232
Contribution of different types of existing geographic information system to formal spatial analysis and to decision making in government.....	240
Overall value of existing geographic information systems .....	251
 Chapter 8 <u>Future Development of Geographic Information Systems...</u>	253
Development of system capabilities assuming present computer characteristics.....	253
Improvements of the non-graphic data base.....	253
Improvements in manipulative capability .....	255
Impact of computer development on geographic information systems .....	259
Development of geographic information systems in context of the overall value of stores of spatial information .....	260

	Page
Future developments within the discipline of geography.....	263
Future developments related to decision making in government .....	267
Institutional data linkages.....	271
Design of formal spatial models .....	273
Post-university transfer of technology to decision makers .....	274
Interaction between data gatherers and data users ...	276
Interaction between data specialists and decision makers .....	277
Summary.....	284
References.....	286
Appendix - Technical description of Canada Geographic Information System.....	306
Published Contribution - A Geographic Information System for Regional Planning.....	following page 444

## LIST OF FIGURES

Figure	Title	Page
1. 1	Relationships between data gathering, data handling, data analysis, and decision making. ....	12
2. 1	Relationships between the spatial dimensions of entities to be recorded, graphic and symbolic records, and the choice of the storage media. ....	32
2. 2	Common location identifiers in existing sets of data. ....	34
2. 3	Variables that influence the character and availability of spatial data gathered by governments. ....	37
2. 4	Relationship between human and instrument capabilities for sensing and calibrating spatial data. ....	44
2. 5	Methods of spatial data sensing and recording. ....	46
2. 6	Relationships between data format, record format, and retrieval possibilities. ....	54
3. 1	Relationship between type of manual technique and time used. ....	78
3. 2	Relationship between data volume and time required for data manipulation and extraction. ....	79
3. 3	Relationships between manual techniques, volume of data, and time used. ....	79
3. 4	Relationship between components of manual techniques and time used. ....	81
4. 1	Developments in data flow. ....	84
4. 2	Components of computer storage. ....	85
4. 3	Data input and retrieval paths. ....	86
4. 4	Relationship between metric and topological codes. ....	88
4. 5	Range of coding related to expression of spatial relationships within a data set. ....	90
4. 6	Operations possible on spatial data coded as coordinate points. ....	96
4. 7	Relationship between real-world data elements and coded non-graphic storage formats. ....	97
4. 8	Types of manipulation capability in systems based on storage formats using arbitrary regular areas as a function of cell size and characteristics of source data elements. ....	98

Figure	Title	Page
5.1	Sample map of 15 irregular polygons. ....	133
5.2	Centroid coordinates (15 data points). ....	133
5.2.1	Fifteen arbitrary regular-area cells. ....	133
5.3	Network representation (32 data points). ....	134
5.3.1	Thirty-two arbitrary regular-area cells. ....	134
5.4	Coarse polygon representation (98 data points). ....	134
5.4.1	Ninety-eight arbitrary regular-area cells. ....	134
5.5	Fine polygon representation; 10 points per in. (330 data points). ....	135
5.5.1	Three hundred and thirty arbitrary regular-area cells. ....	135
5.6	Fine polygon representation; 50 points per in. (1,650 data points). ....	135
5.6.1	One thousand, six hundred and fifty arbitrary regular-area cells. ....	135
5.7	Sample map with single irregular polygon indicated. ....	138
5.7.1	Polygon boundary assigned to regular grid using same data volume (15 data points) as centroid coordinates. ....	138
5.7.2	Polygon boundary assigned to regular grid using same data volume (32 data points) as line network. ....	138
5.7.3	Polygon boundary assigned to regular grid using same data volume (98 data points) as coarse polygon. ....	139
5.7.4	Polygon boundary assigned to regular grid using same data volume (330 data points) as fine polygons at 10 points per in. ....	139
5.7.5	Polygon boundary assigned to regular grid using same data volume (1,650 data points) as fine polygons at 50 points per in. ....	139
5.8	Maximal neighbourhood of a point in a matrix. ....	141
5.9	Representation of an irregular polygon by maximal neighbourhoods. ....	141
5.10	CGIS: overall system concept. ....	153
5.11	Operations involved in CGIS process of graphic to non-graphic conversion. ....	154
5.12	Image Data Set - Level 1. ....	156
5.13	Descriptor Data Set - Level 1. ....	157
5.14	Structured Descriptor Data Set. ....	157

Figure	Title	Page
5. 15	Information loss in relation to sizes of storage and output elements. ....	180
5. 16	Transformations between storage and output formats undertaken by typical systems. ....	181
6. 1	Axes of descriptive framework. ....	188
6. 2	Location identifiers with related data sets. ....	189
6. 3	Categories of volume of related data with examples of typical systems. ....	192
6. 4	Manipulation facilities and their related systems. ....	195
6. 5	Example of the merge process. ....	202
6. 6	Intervisibility between given points. ....	203
6. 7	Overlay condition where Map B is partially overlaid on Map A. ....	205
6. 8	Location identifiers related to volume of related data. ....	211
6. 9	System manipulation facilities related to location identifiers. ....	213
6. 10	System manipulation facilities related to volume of data per location. ....	215
6. 11	Diagonal OS as a vector of potential capability of system file structures. ....	218
7. 1	Relationships between data, analysis, and decision. ....	222
7. 2	Typical set of decision-making institutions. ....	233
7. 3	Typical types of decision within national, regional, and local institutions. ....	233
7. 4	Simplistic concept of policy, management, and administrative decision levels in national, regional, or local government in a parliamentary system. ....	236
7. 5	Categories of geographic information systems, their characteristics, primary utility, and primary use. ....	241
8. 1	Complexities in the relationships between decision making, geographic information systems, and spatial analysis. ....	261
8. 2	Future developments within the discipline of geography. ....	264
8. 3	Future developments of decision making in response to developments in geographic information systems and spatial analysis. ....	267
8. 4	Conceptual sequence of events between data gathering and policy decision making within government. ....	269
8. 5	Interactions between data collection, spatial data handling, mental and formal spatial analysis, and decision making, over time. ....	280



## LIST OF TABLES

Table	Title	Page
3.1	Summary of results of distance estimation and measurement. ....	74
3.2	Summary of results of area measurement test. ....	76
5.1	Number of data points required to define a series of irregular polygons by chain-encoding boundaries and by skeleton method. ....	142
5.2	PIOS: computer time required to process an average map from digitization to Master File format. ....	149
5.3	CGIS test: summary and time and cost analysis. ....	162
5.4	CGIS and manual equivalent times for handling data at various densities. ....	164
5.5	CGIS test: total elapsed time and cost required to create data bank. ....	165
5.6	CGIS test: retrieval costs. ....	167

## CHAPTER 1 - INTRODUCTION

The objectives of the thesis are to examine new developments in techniques for handling spatial data, and to consider the implications of their use on the methodological aspects of geography and on decision making in government.

Data handling fits within a larger framework of data gathering, data analysis, and decision making. Conceptually, the interactions among the elements in this framework are expressed in the following diagram (Fig. 1. 1).

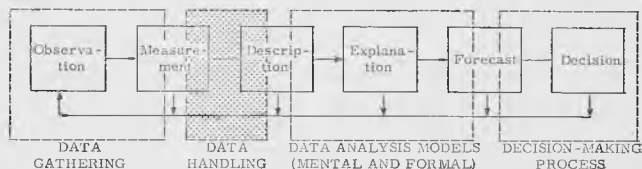


Fig. 1. 1. Relationships between data gathering, data handling, data analysis, and decision making.

The steps necessary to arrive at decisions based on observed phenomena are suggested by the small boxes in the diagram. For the purposes of this discussion it is assumed that phenomena are observed, measured in some suitable unit(s) and described on the basis of these measurements. When the phenomena have been described, an attempt can be made to explain their properties, such as their spatial distribution. If they can be explained, future changes in the variables can be suggested and decisions can be made on the basis of the predicted outcome.

These steps can be grouped into somewhat more familiar categories. Data gathering encompasses the various techniques of census, survey, administrative record keeping, and so forth. Data handling comprises the storage, manipulation, and display of data. Data analysis involves fitting the data to some form of model, either mentally intuitive or more rigorous and formal, that allows the phenomena to be understood and, with greater or lesser degrees of accuracy, predicted. Decision-making processes, whether academic, pragmatic, individual, institutional, or political, can then result.

It can be seen that data handling is a limited part of this process. It is an essential and integral part of the series but it does not encompass data gathering, the design and use of techniques for data analysis, or the decision-making process. This thesis will concentrate on the handling of spatial data, but will relate the handling techniques to the characteristics and volumes of available data and to the subsequent needs of various types of data use and decision making.

Data handling itself is an extremely large field. To limit the scope of the thesis, attention will be focussed on techniques for handling the mapped and related location-specific data gathered by government agencies. The special techniques for handling data from pictures (picture processing) will be ignored. Similarly, only government decision making will be examined. Although no direct examination will be made of the private and commercial sectors, or of the overwhelming majority of data which is not of a location-specific type, an extensive subject nevertheless remains.

Mapped data and related location-specific data have probably been gathered and handled since at least 2000 B. C. Of necessity, manual cartographic techniques were employed to create and read maps, and such techniques are still the most widely used in the world today. Manual techniques have some advantages, but they also have significant disadvantages. Although maps can be an efficient way to store and display data concerning the earth's surface, there are two basic limitations to their use. The first is that there are physical limits on the amount of descriptive data that can be stored and displayed on any map. To reduce such data to a sensible graphic form, they must be classified and generalized, which too often causes a loss of detail during the transition from a less to a more generalized form of data record. The second limitation is that information in a map format, as we know it, has to be retrieved visually and manually. Measurement is laborious, and quantitative comparisons are slow and expensive. Any large collection of cartographic material presents a formidable task of reading and analysis to obtain even the simplest understanding of the information it contains.

The techniques and effort involved in the manual handling of mapped data will be examined in detail in Chapter 3. The subsequent chapters of the

thesis will consider the techniques that have been developed in the past 15 years (since the advent of the electronic computer amenable in the handling of mapped data). Both manual and computer-aided techniques will be brought together and reviewed in Chapter 6, so that the current overall capability for handling spatial data may be understood. This capability will then be viewed in Chapters 7 and 8 in terms of the need to handle data for spatial analysis and decision making in government.

#### REVIEW OF LITERATURE

It is clear that the thesis draws from, and is concerned with, the relationships between cartography, computing, information system design, geographical explanation, formal spatial analysis, and to some extent political science. Each of these fields has an extensive literature but, as may be expected, work in them tends to focus on matters central to each discipline rather than exploring their interrelationships. Even texts that attempt to encompass the scope of their disciplines are weak in this respect.

The classical texts in cartography, such as the work of Robinson or Monkhouse and Wilkinson,<sup>1</sup> are concerned mainly with the "how" of producing maps and diagrams rather than the "why". The overall literature on computing is vast, but that specifically concerned with handling spatial data has been sparse. (It will be more fully examined below and in the body of the thesis.) The concepts of information system design are generally applicable to all data, but the systems used as examples<sup>2</sup> have concentrated on the needs of business management, industrial control, economic planning, military systems, and so forth. The several excellent reviews of formal spatial analysis and geographical explanation<sup>3</sup> do not cover data handling techniques in any depth. The extensive literature of political science, even that directly concerned with decision making in government,<sup>4</sup> does not often explore the supply and handling of data on which decisions are based. Even studies specifically concerned with government management of the environment<sup>5</sup> make only limited reference to the inherent task of handling spatial data.

It would be wrong, however, to suggest that the problems of handling

1. Robinson 1969, Monkhouse and Wilkinson 1970. 2. McDonough 1963, Cleland and King 1969, Dodd 1969, Burch and Strater 1974. 3. Chorley and Haggett 1967, Cole and King 1968, Berry and Marble 1968, Harvey 1969, Rushton et al. 1971. 4. Ward 1962, Porter 1965, Bray 1970. 5. MacNeill 1971.

spatial data, and the possibility of bringing new methods to bear on these problems, have gone unrecognized. Significant contributions have been made to various aspects of the work. The calculating capacity of computers was readily perceived by numerous workers. In particular, their capabilities for storage and information handling, in terms of spatial data manipulation, were anticipated by Coppock, Coppock and Johnson, Kao,<sup>1</sup> and others. As the development of computers progressed, their contribution to spatial data handling was explored and noted by a series of workers. Tobler<sup>2</sup> examined the automatic production of thematic maps. Dacey and Marble<sup>3</sup> and Dueker<sup>4</sup> made early comments on the needs and theoretical problems of handling spatial data with the aid of computers. Pfaltz<sup>5</sup> and Hägerstrand<sup>6</sup> generally considered the computer-aided handling of spatial data in terms of applications to specific geographical problems. Balchin and Coleman,<sup>7</sup> followed by Bertin,<sup>8</sup> produced two of the rare overviews of computer use within a conceptual view of cartography. Haggett<sup>9</sup> presented a timely review of computer use in geography. Aumen<sup>10</sup> considered the advantages of numerical maps.

With few exceptions, however, these general contributions have dealt with the potential of manipulating spatial data with computers, usually with the aim of producing a map. The broader concept of computers as tools to link the processes of spatial data gathering and decision making is not often considered.

Two other categories of relevant literature can be recognized. The first discusses devices or techniques that contribute to the overall capability for handling spatial data, and the second describes the development of sets of techniques into geographic information systems with various characteristics and capabilities.

Existing electronic computers are mainly general-purpose machines for digital data processing. Hardware for handling graphic data is limited to input and output devices, and as they have developed rapidly, the literature usually covers specific types of equipment. Finkelstein<sup>11</sup> analysed input techniques, Boyle<sup>12</sup> and Appel et al.<sup>13</sup> discussed interactive cathode ray

1. Coppock 1962, Coppock and Johnson 1962, Kao 1963. 2. Tobler 1965.
3. Dacey and Marble 1965. 4. Dueker 1966. 5. Pfaltz 1967a. 6. Hägerstrand 1967a. 7. Balchin and Coleman 1967. 8. Bertin 1970. 9. Haggett 1969.
10. Aumen 1970. 11. Finkelstein 1970. 12. Boyle 1970a. 13. Appel et al. 1968.

tubes for data input and editing, and Petrie<sup>1</sup> presented an overview of cartographic plotting devices for automatic mapping. Dueker<sup>2</sup> has published the clearest and most straightforward review of the methods of coding geographical patterns, though other workers<sup>3</sup> have provided alternative schemes. Williams<sup>4</sup> produced a valuable bibliography of work on data structures and later specifically reviewed those of use in computer graphic systems.

Techniques for specific manipulations of spatial data in numerical form have been developed by many workers. Boehm<sup>5</sup> compared various techniques of surface sampling and concluded that there is no optimal technique. The task of drawing contours has received a great deal of attention; probably the best discussions are the early one of Bengtsson and Nordbeck<sup>6</sup> and that included in the work of Harbaugh and Merriam.<sup>7</sup> Specific contouring programs have been developed by numerous workers,<sup>8</sup> and techniques of trend surface analysis received early attention from geologists, notably Krumbein.<sup>9</sup> Chorley and Haggett,<sup>10</sup> Harbaugh and Merriam,<sup>11</sup> and Cassetti and Semple<sup>12</sup> have reviewed the work of others. The computer-aided manipulation of map projections has been a continuing interest of Tobler<sup>13</sup> and Kao.<sup>14</sup> Map generalization was discussed by Tobler<sup>15</sup> and Koeman and van der Weiden,<sup>16</sup> among others.

This sample of publications confirms that a considerable amount of work has been undertaken on specific aspects of spatial data handling.

Tarrant<sup>17</sup> has provided a partial list of the computer programs available in the geography departments of universities, and Peucker<sup>18</sup> has compiled a bibliography on computer cartography.

Literature dealing explicitly with geographic information systems is less readily available in established journals. It is mainly to be found in the form of government publications, conference proceedings, system-related manuals, or mimeographed reports. Various categories of geographic

1. Petrie 1966. 2. Dueker 1972. 3. Freeman 1961, Pfaltz and Rosenfeld 1967, Jarvis 1971. 4. Williams, R. 1968, 1971. 5. Boehm 1967. 6. Bengtsson and Nordbeck 1964. 7. Harbaugh and Merriam 1967. 8. Murray 1968, Dayhoff 1963, Hebin 1969. 9. Tobler 1970, Holroyd and Bhattachayya 1970. 10. Krumbein 1959, 1966. 11. Chorley and Haggett 1965. 12. Harbaugh and Merriam 1967. 13. Cassetti and Semple 1968. 14. Tobler 1962. 15. Kao 1963. 16. Tobler 1966. 17. Koeman and van der Weiden 1970. 18. Tarrant 1968 and subsequent editions. 19. Peucker 1972.

information systems will be described in the body of this thesis, with specific examples. A bibliography of the literature is provided.

#### CONTRIBUTION AND CONTEXT OF THESIS

The contribution made by this thesis is to provide a coherent examination of the tools now available to link the processes of spatial data gathering and decision making. Manual and computer-aided techniques for handling spatial data are viewed not in terms of their ability to produce maps, but in their broader role of storing, manipulating, and displaying data for decision-making purposes. Particular attention is paid to the utility of the recently developed non-graphic forms of data storage as a repository for spatial data. The implications of the use of the new tools for formal spatial analysis and government decision making are discussed and guidelines for the future development of techniques for handling spatial data are put forward.

The tools for handling spatial data must of course be thought of in the context of the intended use of the data. The supply of data and the demand for them are not independent. In general, data are obtained in response to a variety of institutional needs, whether historical, current, anticipated, real, or illusory. This influence of the source agencies is very clear when one looks at the supplies of mapped and related location-specific data. Available data reflect the institutional requirements within a particular country, which in turn reflect their different needs and states of development. Although this situation may mean that data are collected for one purpose only and cannot be transferred to other uses, it does generate data that at least have some use. Data have no value except in the context of their use, and in the past they have not been generally collected for their own sake.<sup>1</sup> The tools for handling data also have not been developed for their own sake, but rather have grown pragmatically out of institutional need. The influence of their institutional origin is particularly pervasive in the current early stage of development of computer-aided techniques for handling spatial data; each of the categories of technique to be examined in later chapters is related to specific types of institutional activity. The techniques are not necessarily applicable in every institution and every country, and no universal system for handling mapped data exists at the

1. However, we may be acquiring such an overwhelming ability to gather data that there is danger of undirected accumulation in the future.

moment or is likely to be created in the near future.

The development of data handling methods, then, has grown from the institutional need to handle data; consequently, their usefulness and applicability in each country must be judged within that context.

#### THESIS STRUCTURE

The thesis starts with a brief examination of the characteristics and availability of spatial data, proceeds with an examination of manual and computer-aided spatial data handling techniques, and ends with the discussion of the relationships between geographic information systems, formal spatial analysis, and decision making in government.

The type and amount of spatial data that exist obviously affect the need for data handling capability. Location-specific data themselves result from the description of entities or conditions within space, and the nature of the description depends on the concept of space and the language employed. The range of these possibilities is briefly examined in the first part of Chapter 2. The influence of the storage medium on the data format is considered. Despite the wide range of conceptually possible approaches to spatial description, the overwhelming amount of existing location-specific data assumes Euclidean space and is stored on documents in either graphic or symbolic data formats. The most commonly used types of location identifier are identified and discussed.

The next step is an examination of variations in the supply of such spatial data. An inventory of spatial data supplies in various countries is outside the scope of the thesis, but the factors that influence data gathering and the range of variation within each factor are discussed in the second part of Chapter 2. It is seen that each data set within each country (or even in different parts of a country) may exhibit characteristics that require different subsequent handling. The overall supply and availability of spatial data may vary substantially from one country to another. Consequently, the usefulness of various approaches to data handling varies between countries and the appropriate mixture of techniques for any one country depends on the data available within it. These observations are made at the outset of the thesis so that electronic computer methods for spatial



data handling can be examined subsequently in a clearer perspective. This preliminary discussion does, however, make it clear that capabilities for data gathering are rapidly expanding. The increased perceived need for spatial data and the growth of the technologies basic to data acquisition suggest that the present rapid increase in volume of spatial data will continue. This growth, whether or not it is well directed, will generate increasing quantities of spatial data which will need to be handled.

Manual techniques for handling mapped data and other location-specific data form the benchmarks of traditional capability against which machine methods must be judged; they are examined in Chapter 3. The efficiency of the map as a source of data from which useful information must be extracted and the techniques for extracting it are discussed. Manual processes are thus reviewed in the context of data handling rather than simply as traditional cartographic practice. Maps are regarded as groups of records, or "files", in which spatial data are stored. The arrangement of data on a map is considered as a "file structure", and manual handling techniques are those used to manipulate, group, and extract data from the files.

A very wide range of spatial data handling tasks can be accomplished manually, constrained mainly by the time and effort needed to perform them; the discussion in Chapter 3 focusses on their efficiency. In general, tasks that rely on manual motor processes, such as data transfer by tracing, or manual measurement, are seen to be slow. Techniques that rely mainly on mental data processing are fast. Techniques that rely on combinations of these processes increase in efficiency as the proportion of mental data processing becomes greater, until the maximum rate at which bits of information can be transmitted to the brain (human sensory channel capacity) is reached. More specifically, maps can be economically handled by manual techniques if the processes involve reading and manipulation or grouping, or both, of data written directly on the map. Graphic source data cannot be handled so efficiently, however, when the data to be extracted by manual measurement and comparison are inherent in their file structure. Increasing data volume rapidly makes it uneconomical to manipulate and

extract even the data actually written in the file. Also, improvements in the flexibility of a graphic file usually imply creation of multiple graphic records, and hence increase the bits of data to be handled; graphic source material can be said to impose an inherent inflexibility on file structures, and is not very efficient as a source of large volumes of spatial data.

The examination of electronic computer-aided techniques for handling spatial data thus proceeds from the observation that the volume of spatial data available in the world is rapidly increasing and that graphic storage media and manual techniques are inadequate to handle large volumes of data.

The storage of spatial data by electronic computers has two inherent physical advantages. The first is the relatively compact nature of the storage compared with conventional graphic storage. The second advantage is more crucial, since it adds a new dimension to the first; this is the ability to use established computing machines to read, store, and manipulate non-graphic stores of spatial data. In addition, such stores of data are relatively easy to augment, edit, and correct without renewing the entire storage. For all computers in general use in 1973-74, however, mapped data and related location-specific data must be available in a coded, machine-readable, non-graphic form of storage.

After the range of coded, non-graphic spatial data formats and the basic attributes of file structures have been outlined in Chapter 4, general format-related categories of computer-aided techniques are examined, and their characteristics are illustrated with specific examples in Chapters 4 and 5. The Canada Geographic Information System, whose development was initiated and directed by the author, is examined within this context. Computer-aided techniques are compared with manual ones, and trends in the development of digitization techniques, interactive data-base manipulation, output devices, and multifformat systems are also examined so that their potential contribution to spatial data handling can be considered.

From this examination comes the view that it is generally cheaper and quicker to use a computer to manipulate spatial data than to handle numerous maps manually, particularly when measurements have to be made and when comparisons between different maps are needed. There are difficulties,

however. The main shortcoming of the computer-aided approach is that currently it is uneconomical to transform the data from maps into a machine-readable format, and this imposes constraints on the amount of data stored in the computer and the way they are stored. Therein lies the challenge and prospect of the new techniques. If it is possible to remove or reduce the difficulties of making maps readable by machine, and there is every evidence that it is, then there is a good prospect of using computers to read and manipulate maps.

Paradoxically, the full capabilities of modern computers are significantly under-utilized by even the most advanced of the existing systems for handling spatial data. This is partly because systems are currently at an early stage of development, but partly also because of a lack of any clearly stated requirements to manipulate the data in ways that use modern statistical or model-based approaches and so take advantage of the computers' capabilities. There are several reasons for this deficiency, some stemming from the type of institutions that have fostered the growth of the new techniques, and some arising from the present methods of decision making and the training of decision makers. Several of the computer-aided techniques, in fact, can be recognized as conceptual extensions of traditional manual cartography, designed to present spatial data in a form that human beings can comprehend by visual examination. Only recent developments in computer-aided techniques have tried to extend human analytical capabilities by systematically comparing and manipulating the spatial data within the computer and presenting the results to the user.

These thoughts are synthesized in Chapter 6. The various sets of techniques analysed in the previous chapters are brought together in an attempt to see what progress is being made, and a tentative framework is provided within which both manual and computer-aided systems are located so that their relative capabilities for handling spatial data can be understood. The contribution of the various techniques to the overall ability to handle spatial data is thus summarized. The categories of different capabilities emerge, their institutional backgrounds are made clear, and several important

questions are raised about the way in which the techniques are developing, their utility, and the likely future progress of geographic information systems. These problems relate on the one hand to geographical theory and formal spatial analysis, and on the other to decision-making processes.

Chapters 7 and 8 attempt to tackle some of these issues, which are obviously complex and not amenable to easy solution. In Chapter 7, the contributions of the categories of existing systems to formal spatial analysis and governmental decision making are examined.

In terms of formal spatial analysis, the new techniques have not contributed to theory creation, but have the potential to speed the course of theory testing and allow new avenues to be explored more rapidly. Specifically, they may help by making available information that would:

- 1) allow current models to be tested over a wider area, or for more disaggregated populations;
- 2) allow the development of new models (often derived from old theory) that may incorporate variables not before available in quantitative form; this may, in turn, lead to completely new types of theoretical development;
- 3) as a special case of (2) above, provide data on changes that occur at particular places, and the interrelationship between changes at places (that is, the interaction between attribute and time over space); this may help to develop and improve dynamic models of innovation, diffusion, the lagged effects of economic and social change, and so forth.

In terms of government decision making, existing geographic information systems are seen to make their main contribution to the mental decision process at management and policy levels of decision making.

In general, it is apparent that although the processes of spatial data handling, formal spatial analysis, and decision making are interdependent, there is a substantive lack of interaction between them. Although the development of computer-aided systems so far has grown out of a combination of pragmatic need and intellectual curiosity as much as from any narrowly defined progression of technical evolution, there have been few linkages

between the development of such systems and the processes of formal spatial analysis or government decision making.

The future development of techniques for handling spatial data is considered in Chapter 8, first from the point of view of the technical aspects of the systems themselves, and secondly, in terms of their relationship with the discipline of geography and the process of decision making in government.

The guidelines proposed for development of techniques are relatively straightforward. Improvement in the techniques for transformation from graphic to non-graphic format is important if digital computers are to be used. The stores of available spatial data will increase if existing non-graphic data sets are made more compatible and quantitative data gathering processes are used more frequently. Development of the capability for manipulation will stem from better use of interactive (man-machine) computer capabilities, better statistical understanding of the manipulation processes themselves, the growth of multimodal systems, and the design of file structures based on further development of languages used for ordering spatially distributed data.

The development of geographic information systems within the context of formal spatial analysis and government decision making is less straightforward. It is not possible to forecast the impact of geographic information systems on the discipline of geography, but it is reasonable to expect some mutual shifts of emphasis in both formal spatial models and in the systems themselves as their development continues. Together with examples of related developments in other disciplines, the evidence suggests the following developments:

- 1) Better use of the a priori knowledge of individual situations will lead to the need to use disaggregated data sets. It also implies the theoretical development of linkages between microdata and macrodata, if understanding of the larger socio-environmental mechanisms is to be gained.
- 2) These developments in turn rest on advances in geographic information systems, both to make better use of available stores of information

and to handle the attribute-space-time interactions among data. This will depend on substantial improvements in the file structures of geographic information systems.

- 3) The development of more efficient file structures will depend to a considerable extent on the theoretical development of multidimensional languages.

This type of mutual development of formal spatial analysis and geographic information systems will have considerable impact on the availability of relevant information to provide understanding of spatial distribution and interactions.

Consideration of the future development of geographic information systems for government decision-making purposes similarly leads to the conclusion that both theoretical and pragmatic advances must rely on exploration and recognition of the interdependence of the subjects. The implied increases in communication will not be achieved without effort. The development of inter-institutional data linkages will in all probability have considerable impact on the supply of available spatial data and hence the type of decision making that can benefit from improved handling techniques and improved formal spatial analysis. The lack of formal training of present-day decision makers in quantitative methods undoubtedly affects the perceived need for and understanding of formal spatial analysis, and hence the design and use of geographic information systems to provide data for both formal and mental models. As an interim measure, it is suggested that those concerned with the development of formal spatial analysis techniques, and in particular with the design of geographic information systems, must understand that it is necessary to encourage users to become competent in quantitative methods of problem solving as well as to build information systems and formal analytical capabilities.

It is recognized that increased capability for gathering data may lead in the foreseeable future to the collection of large volumes of data, which we may or may not be able to handle and which we may or may not need. It is apparent that increased feedback from policy levels of government is needed to identify which spatial data should profitably be collected, and to provide the criteria for delimitation. As far as is known, no organizational

mechanisms are in place to improve that feedback. It is suggested that the first improvements in feedback may come from the use of geographic information systems. Only when data are actually read and used in mental or formal models, and it can be shown that changes of content or format are necessary or that other data might be profitably collected with the same expenditure of funds, can persuasion be brought to bear on traditional data gathering agencies to amend their standard forms of operation.

The second level of improvement in feedback will come when improved spatial models are integrated into the decision-making process and can be used to specify data requirements. This implies a change in the decision-making process as we know it today. A substantial degree of integration of formal techniques in the decision-making process would be needed before their data needs would represent a strong argument for changing established data gathering practice.

The possibility of such changes in the decision-making process is discussed in Chapter 8 and possible organizational structures are postulated. In addition, the recent development of interactions between data collection, spatial data handling, mental and formal spatial analysis, and decision making are reviewed and the timing of future changes is a subject of speculation. It is seen that the development of geographic information systems cannot sensibly proceed further in isolation, but must be undertaken as an integral part of a very large structure of data gathering, data analysis, and decision making.

## CHAPTER 2 - CHARACTERISTICS AND AVAILABILITY OF LOCATION-SPECIFIC DATA

The objectives of this chapter are twofold: first, to outline the characteristics of location-specific data that affect their subsequent handling, by either manual or computer-aided procedures; and secondly, to give an understanding of the factors that affect the generation of such data and determine their supply and availability.

The data of concern in this thesis are those derived by observation and measurement of certain aspects of the earth's surface, the environment, resources, population, and related attributes within the living space of mankind.<sup>1</sup> These data are often displayed on maps produced by government agencies. To limit the scope of the discussion, the location-specific data referred to will be those generally gathered by government agencies.

### CHARACTERISTICS OF LOCATION-SPECIFIC DATA

The first task is to define the meaning of "location-specific" data. The information content of a single datum, or "data element", can be regarded as having three potential components. These are "substance" information, "locational" information, and "temporal" information. Substance information describes the attributes, characteristics, variables, values, features, and similar qualities of the datum; the range for any one datum may obviously be multidimensional. Locational information describes the position of the datum in space relative to other data. Temporal information describes the instant or period in time for which the datum is valid. Figures concerning the accuracy of any of the three potential components may be added as supplementary information.

One or more of the information components is present in each datum. If the locational information is present, the datum is termed "location-specific". A location-specific data set is a collection of data that are individually or collectively attached to location; "spatial data" is a term used synonymously with location-specific data. Spatial data systems or

1. No rigorous definition of the living space of mankind will be attempted. In general terms it will be taken as the space contained on the earth's shell between the Moho discontinuity in the crust and the upper limit of the atmosphere, in that man as an unsupported organism must still (1974) remain within those bounds. The upper limit, in particular, however, can be transcended by mankind with the aid of vehicles.



geographic information systems are sets of techniques that allow the locational information of the data to be manipulated in concert with their remaining value.

To provide locational information it is necessary to have a concept of space and a language in which to express a description.

A full discussion of the concepts of space, the structure and capabilities of available languages, and the methods or "models" of measurement and classification is outside the limits of this thesis. The comments that follow merely outline the assumptions that underlie current practice, so that the characteristics of existing spatial data can be better understood and the scope for future developments can be visualized.

#### Concepts of Space

Concepts of space make a distinction between physical and perceptual truth. Physical truth rests upon experiment. Within the limits of current experiments, most writers agree (though this is still an open question) that physical space for practical purposes is Euclidean in structure; space can generally be thought of as having three dimensions. Time may thus be considered to be a fourth dimension. (Where extreme extraterrestrial distances are concerned, matter, or fields, may be considered as a fifth dimension, though this has no practical application in specifying the relative location of data related to the earth's surface.)

Perceptual space,<sup>1</sup> on the other hand, may have numerous dimensions. Relationships between objects on the earth's surface, the extent of areal units, and so forth, can be measured only in four dimensions or less, in practical terms of physical space. However, in perceptual terms, they may involve many "dimensions", such as cost, satisfaction, utility, opportunity, social interaction, and so on. An interesting illustration of this concept is Isard's<sup>2</sup> perceptual view of "political time" as a path traced in a conceptual three-dimensional framework of calendar time, time to the next election, and lifespan time of the politician. Such a multidimensional perceptual view of time may be required to understand government decision making and to describe it adequately. Similarly, a view of distances, extent of areal units, and so forth within a multidimensional

1. The representation of space by means of imagery concepts. 2. Isard 1970.

perceptual view of space may be necessary to understand the distributions and interactions of objects on the earth's surface. The concept of space employed, as well as the empirical properties of the concepts being examined, underlie the choice of language used.

#### Spatial Languages

Items that occupy positions in space can be named. The replacement of proper names by the terms of some coordinate system allows the position of the items to be more generally stated. The sets of rules governing the use of such coordinate systems can be regarded as "spatial" languages. The Euclidean concept of space underlies Euclidean geometry, which has dominated the descriptions of space for over 2,000 years. Its simplicity is its strength, and many geographical concepts are cast in Euclidean terms simply to allow use of the analytical tools of Euclidean geometry. In the 17th century, complex (Euclidean) geometrical problems were solved by algebraic means. Vector and tensor algebras were later developed to handle two- and three-dimensional geometrical problems. Geometry can similarly be related to probability theory, number theory, lattice theory, and so on. The significance of the algebraic approach is that geometric concepts and theorems can be extended to three, four, and up to an infinite number of dimensions. The 19th century, however, saw the emergence of non-Euclidean geometries, which were based on different concepts of physical space. Harvey<sup>1</sup> notes that:

"The hyperbolic space of Lobachevsky and Bolyai was the space of constant negative curvature, the elliptic space of Riemann was the space of constant positive curvature, and the flat space of Euclid was the space of zero curvature. Every perceptual shape from a flat board to a Henry Moore statue could be given a coordinate system, which would define the intrinsic geometry of that shape... so general was the theory that Riemann devised that it could be extended to more than three dimensions."

Geometries thus exist that allow discussion of n-dimensional space. At the end of the 19th century, Klein<sup>2</sup> in particular developed a new geometry founded on the topological characteristics of objects. This geometry is essentially nonmetrical, and concerned only with the continuous connectedness of points on a figure.

A considerable range of languages is thus available for spatial description.

1. Harvey 1969. 2. Klein (English ed.) 1939.

for the assignment of location identifiers to data. Topology, for example, is directly applicable to the study of the connections between settlements by a set of transportation links. Calculations relating to curved space, such as transformations from one map projection to another, can efficiently make use of the more recently developed analytical geometries. The calculation of angles and distances in navigational practice, on the other hand, is much simpler by Euclidean than spherical geometry. Algebras provide convenient ways of handling calculations in multidimensional space.

The spatial languages, however, have been developed in disciplines other than geography. Their application to geographical problems is growing, but represents a future potential rather than a widespread current practice. This particularly applies to the gathering of spatial data. The spatial language of latitude and longitude (or local Cartesian coordinate systems), based on Euclidean space, is used for almost all maps and location identifiers available today.<sup>1</sup> Euclidean geometry is the tool most frequently used for deriving measurements and comparisons from such data. In subsequent chapters it will be seen that the file structures of spatial data systems are also influenced by this tradition rather than by the concept of perceived space or the languages amenable to multidimensional calculation.

#### Spatial Description

Description of items may be nominal or ordinal. In nominal description (binary measurement), the object or event is given a proper name (label), or one relating to a system of classification. The connection between such a name and a location identifier may be established by reference to a separate description, but is not inherent in the name itself. Ordinal description implies some form of measurement. Measurement is not desirable or undesirable in itself; it is merely useful or not useful. It does, however, presuppose a concept of space and a suitable language. The scales used for measurement may be single (scalar measurement), such as ordinal, interval, or ratio scales, which assume a single, one-dimensional, underlying continuum; or multidimensional (vector, tensor, or other measurements). Location identifiers presuppose multidimensional measurement and, for Euclidean concepts, assume orthogonal axes that possess the same scale characteristics. Measurement in perceptual space, however, may use

1. A notable exception is the data gathered by the ERTS satellite sensing devices. See Chapter 5, p. 178.

axes that are not orthogonal and may mix axes that have ordinal, interval, or ratio scales.

The empirical properties of the objects or events being described, and the need for the description, obviously have impact on the characteristics of the data (including the location identifier).

If objects or events are represented as a series of separate points in space, in time, or in space-time, they are thought of as "discrete" items. If they can be represented as lengths,<sup>1</sup> areas, or volumes in space or in space-time (movement), they may be thought of as "continuous" items.<sup>2</sup> Both discrete and continuous objects or events may be countable, finite but uncountable,<sup>3</sup> or infinite in number. Only if they are countable by available techniques of observation and measurement can a "complete" observation be made. In all other cases, and even in many instances where they are countable, a sample is taken. The way in which this sample is taken obviously has impact on the use of the data. A sample may be designed to represent continuous data by discrete data, or vice versa.

#### Spatial Data Format

Data format is defined as the way in which data elements are represented and stored in records. Groups of records are termed "files", and the arrangement of data within a file is the "file structure". Records are physical entities and the way in which data are stored in them depends partly on the language employed and partly on the characteristics of the storage medium. Storage media may be essentially one-dimensional (such as rulers, pieces of string, single-track magnetic tapes), two-dimensional (documents, photographs, computer core storage), or three-dimensional (globes, hardware models, stereophonic gramophone records, hydrographic models, holograms,<sup>4</sup> etc.).

Within any spatial concept, it is possible to envisage discrete points, con-

1. Lengths in one-dimensional time are represented by a number of units and hence are discrete. Perceptual time such as Isard's "political time" is a length through a perceptual space. 2. Certain events may have both discrete and continuous components, i.e., those involving intermittent motion. 3. For example, insect populations of a country, sand grains on a beach, etc. 4. Holograms are a special case of three-dimensional storage in that the three-dimensional data are stored without transformation on molecules within a two-dimensional medium but are extracted, separated, and recombined to form a three-dimensional holographic model using coherent light.

tinuous lines, areas, and volumes (items which in themselves have no dimension, one dimension, two dimensions, or three or more dimensions). These items may be described in a variety of spatial languages, either in symbolic form, using alphanumeric and special symbols, or in graphic form by diagrams. If the language is expressed in symbolic form, it can be recorded on any of the storage media mentioned above. If the language is expressed in graphic form, entities with dimensions the same as or less than those inherent in the chosen storage medium can be described without transformation and the information reduction that the transformation may incur.

If two-dimensional documents are used as a storage medium, the concept of Euclidean space is adopted, and a graphic format is employed, only points, lines, and plane surfaces can be adequately represented (Fig. 2. 1). Warped planes<sup>1</sup> and volumes must be transformed before being recorded. The same media constraint applies when multidimensional space is assumed. In dealing with four-dimensional space-time, for example, a two-dimensional medium can adequately contain information on linear distance  $x$  time (that is, a graph of velocity) as a single record. However, surface  $x$  time or other constructs of three or more dimensions cannot be represented without transformation.

Similarly, the graphic description of the interactions between objects or events, or both, is limited by the dimensional character of the storage medium. Specifically, the highest number of dimensions possessed by any one entity involved in the interaction must not exceed that of the medium. On two-dimensional storage, for example, it is possible to record points in a set of points (for example, brick buildings among wooden buildings), points on lines (railway stations), lines on lines (unpaved parts of a highway), lines in areas (trails in a park), or areas in areas (islands in a lake), but it is not possible to show the location of points, lines, planes, or volumes within a volume without transformation (submarines in the sea, rocket trajectories, cold fronts in an air mass, or clouds in the sky).

1. Warped planes (for example, an irregular topographic surface) can be considered to be a form of "extended" two-dimensional feature. They are frequently recorded on maps by use of point patterns, or isolines (without which it would be considerably more difficult to represent volumes). A transformation has, however, occurred from the warped plane to its representation by points or isolines.

SPATIAL DIMENSIONS OF ENTITIES  
TO BE RECORDED

		0	1	2	3	4	5 → 5+
		Equivalent in Euclidean geometry					
		point	line	surface (area)	solid (vol.)		
		Equivalent in topology					
		node (vertex)	arc (edge)	region (face)	3-D fig.		
DIMENSIONS OF STORAGE MEDIA	1	G	G	S	S	S	S →
	2	G	G	G	S	S	S →
	3	G	G	G	G	S	S →
GROUPED RECORDS		G	G	G	G	(G)	S →
		S	S	S	S	S	

- G = Graphic expression of spatial language possible  
 S = Symbolic expression of spatial language possible  
 (G) = Limited graphic expression of spatial language possible  
 □ = Overwhelming majority of existing spatial data records

Fig. 2.1. Relationships between the spatial dimensions of entities to be recorded, graphic and symbolic records, and the choice of the storage media.

Despite the limitations, however, the overwhelming majority of existing records are stored on two-dimensional documents. Nearly all spatial data are (in 1974) initially portrayed on maps, diagrams, or photographs. The process of transforming multidimensional data to a two-dimensional graphic format underlies the whole tradition of map projections, cartographic symbolization, and the more recent transformations of perceptual space to graphic form.

It was mentioned above that a complete observation can be made only where objects or events are countable. For finite but uncountable, or infinite, numbers of objects (and frequently for objects within countable sets), a sample is taken. The constraint of measurement on the sampling technique

may be that multidimensional objects are represented by data with fewer dimensions. This is the case when point data at weather stations are used to sample the continuous volume of the atmosphere, drill cores are used to sample the inside of ore bodies, and aerial photographs are used to sample topography. Although the data from such measurements can be recorded on a storage medium that does not adequately represent the entity being sampled, a mental or calculative transformation is required to understand the entity from the recorded data.

Records in graphic format can be grouped, in an attempt to overcome the dimensional constraint of the medium. Many pieces of one-dimensional string could be laid side by side on a table top and a two-dimensional map could be drawn across them. A piece of string could be wrapped around a drum and the outline of a three-dimensional figure could be traced upon it. Two-dimensional documents can be superimposed, each representing a "slice" through a volume. (This is done daily with meteorological charts of various levels of the atmosphere.) Three-dimensional hardware models can be caused to move to represent four-dimensional space-time. The ability with which such "files" can be constructed and handled is clearly a constraint of the media.

Recent developments in the field have led to greater possibilities. Two-dimensional photographs can be shown in sequence quickly enough to add a perceptual dimension of time (cinema). Stereoscopic pairs of photographs can be viewed simultaneously and a perceptual three-dimensional surface can be observed and measured (photogrammetry). Records from one-dimensional magnetic tape can quickly generate two-dimensional graphic arrays, and if the arrays are produced in fast sequence a perceptual time dimension is added (television). If the arrays are written on to two-dimensional computer core storage in rapid sequence, a limited three-dimensional representation of space is generated in which calculations may be performed. The future development of computers with parallel processing capability (multiple two-dimensional arrays), or holographic memories, would allow three-dimensional storage to be rapidly manipulated and four or more dimensions to be simulated for measurement.

Clearly, the storage medium affects the data formats. In fact, even from this preliminary discussion it is evident that the empirical properties of the entity or condition being examined, the concept of space employed, the language used, the method of description, the data format, the storage medium chosen, the file structure chosen, and the capability for manipulating the data are all interdependent.

Within this continuum, it should be noted that the Euclidean concept of space, Euclidean geometry, and the widespread use of two-dimensional documents for 2,000 years have had a strong influence on the types of spatial data that actually exist and have to be handled today. The second part of this chapter will examine the factors that affect the supply and availability of spatial data and the following chapters will examine traditional and new techniques for their handling and manipulation.

#### Types of Location Identifier

To summarize the characteristics of existing data, it can be said that five main types of location identifiers are currently in use (Fig. 2.2).

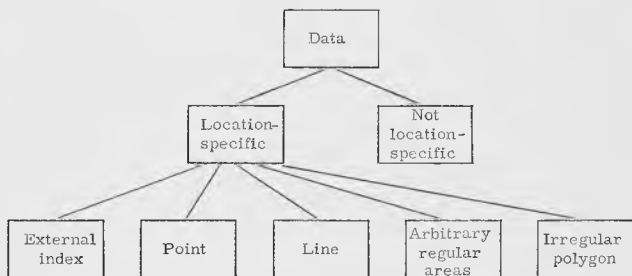


Fig. 2.2. Common location identifiers in existing sets of data.

- 1) The external index comprises nominal labels that are generally descriptive of a location, such as the names of administrative areas, street addresses, census tracts, postal codes, and so forth. Such labels can be used to group data that have no other location identifier. It is, however, impossible to tell the relative position or physical



limits of the elements without reference to a master index (typically a map) showing the boundaries of the named items. Most non-mapped data with any location identifier today are recorded by this descriptive method.

All the remaining types of location identifier are related either to some coordinate system or to other data elements by one or more spatial languages (see discussion above).

- 2) Discrete points are best employed as abstractions of small phenomena, but frequently are used as surrogates for larger phenomena. The centroid of an irregular area, for example, may be used as the location at which the characteristics of the area are recorded. This approach, unlike the use of nominal codes, allows the relative positions of the larger phenomena to be established, but it does not, of course, define their boundaries. Sets of points may also be used to represent entities with one or more dimensions (for example, spot heights on a topographic surface, or weather-station data in an air mass).
- 3) Data can be assigned to line segments. Commonly, line segments are employed to represent features such as roads, railways, flight paths, geological faults, and streams. Line segments are usually plotted on maps and thus their relative positions are recorded within a coordinate system. They may, however, be topologically related and graphically displayed without reference to coordinates, as on a strip map of a route or an underground map of London. Line segments are one-dimensional, but a network of such segments is a two-dimensional entity.
- 4) Arbitrary regular areas are used to provide cells for the local subdivision or grouping of spatial data. Their utility depends on the size of the cell in relation to the phenomena being recorded. They are necessarily arbitrary units, typically a grid. The position of the grid in Euclidean space can be specified by its corner coordinates and the dimensions of the grid. The regular arrangement of the cells, the ease of establishing markers on the ground to delimit the regular pattern, and the ease of locating a specific cell within the set led to their early adoption for the subdivision of property (for example, in

Roman cities or North American prairie farms), for purposes of military and administrative reference, and, more recently, for the grouping of data in preparation for computer storage. Criticism of the approach for grouping centres on the fact that it imposes an unreal data format on the real-world phenomena.

- 5) Irregular polygons are surface-descriptive location identifiers. They differ from line segments in that their boundaries separate areas with unlike characteristics, whereas a line segment or line network may traverse a homogeneous area. The attributes of a surface can be expressed as continuous, single-valued functions, or as step functions of independent continuous variables. These are commonly treated as either contours or sets of labelled regional boundaries. As regional boundaries, irregular polygons frequently define areas grouping data that are otherwise not location-specific (such as administrative areas, postal zones, or census tracts). Used as isolines, irregular polygons extend the capabilities of the two-dimensional document format in that they present a useful approximation of a warped plane; topographic contours are probably the most readily understood example.

These location identifiers can all be graphically recorded on a two-dimensional medium. The graphic expression of the spatial language can be an extremely compact form of notation. The information content of a line can be many times higher than, for example, a numerical expression that uses the same amount of space on a document.

Any of the graphic forms of location identifiers mentioned above can be mixed to form a composite description of a surface. They are, in fact, the stuff that maps are made of, and maps have served man well and continue to do so. A strong argument can be made that graphic representation of spatial data follows man's perceptual understanding of space, and thus is easy to read and assimilate. The ease with which man can extract data from the graphic format will be examined in Chapter 3.

All the above types of location identifier will be examined in greater detail in the following chapters of the thesis. As has been pointed out, they do not represent all possible categories of location identifier; they may not even be the most useful way to specify location in all instances. They do, however,

reflect the concept of space, geometry, and type of graphic portrayal that have found general acceptance and underlie the main body of existing location-specific data. The remaining part of this chapter will examine the factors that influence the overall supply and availability of location-specific data.

#### SUPPLY AND AVAILABILITY OF DATA

The state of development of spatial data handling techniques will be examined later, but it should be made clear at the outset that the usefulness of specific techniques depends to a considerable extent on the supply of data that must be handled. This may vary substantially from country to country. It is not possible within the scope of this thesis to examine the data supplies generated by various countries. A brief examination will, however, be made of the factors that influence the provision of spatial data, so that the existing range of variation can be better understood.

The conditions and circumstances that underlie data supply are complex and interdependent. For convenience, however, they can be grouped into six broad categories: perceived need for data; constraints on data gathering; constraints of capability for gathering; constraints on validity; physical handling characteristics; and constraints on use. The categories can be regarded as variables that are closely interrelated and together determine spatial data availability. They are illustrated in the diagram (Fig. 2.3).

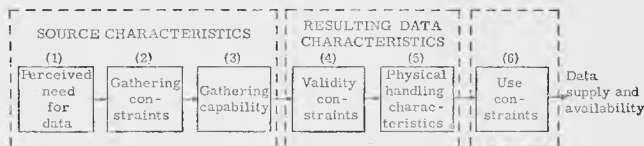


Fig. 2.3. Variables that influence the character and availability of spatial data gathered by governments.

#### Source Characteristics

Perceived need for data. The spatial data for which a government has a "perceived need" are quite simply the data that a government thinks it needs and is prepared to pay for. They are related to the type and level of

economic development of the country, its political and social activities, and, as a special part of these, its military history and aspirations. The total amount of existing spatial data in a country reflects its historical sequence of perceived needs and, to a lesser extent, its view of the future.

An obvious difficulty in proceeding rigorously with a discussion of needs is the lack of any widely accepted scale of economic, political, social, or military development, alone or in combination. A continuum of some kind may exist, as a path traced in a complex, multidimensional perceptual space. The relationship between such a scale and existing spatial data would probably not be linear. It is thus practical only to indicate briefly the range of possibilities and to comment on their relationship to data supply.

Levels of economic development among nations can conveniently be thought to range from dominantly subsistence economies, through primary economies with extractive industries and specialized food production, to capital-intensive industrial economies, and technologically intensive economies. Commercialization in the economy generally increases as overall reliance on subsistence food production decreases.

Somewhat simplistically, it could be suggested that the primary instincts that motivate a government to gather spatial data are territorial control and taxation. At the lower end of the economic scale, spatial data are required for the taxation of land and for the establishment of territorial boundaries within which taxes may be levied. The next step related to economic development is the acquisition of data concerning the exploitable resources of the land, its climate, vegetation, soil, geology, water, and so forth. Initially, these are usually hurried exploratory surveys, with a high subjective content used to direct the investment of effort (capital). Usually, however, they lay the foundations for continuous data gathering in their respective disciplines.

The initial growth of commercialization brings with it the need for spatial data on transportation routes (which may in some instances precede territorial investigation of any kind). The first need is for data from simple traverses, then topographic data for road, rail, and canal building, and for water and air navigation. As population density increases, the need for a

more detailed specification of the tax base arises. Formal population census, agricultural census, and commercial census by taxable area are generated.

The growth of capital-intensive industries (and a new tax base) increases the need for a census of industry. As industry makes demands on national and human resources, and control over their exploitation is deemed prudent, more detailed surveys are required. Government-controlled surveys of logging and reafforestation (as opposed to timber resource surveys) illustrate this need.

As technological development and social living standards increase, there is a need for data on which to base research. Spatial data on forest pests, crop diseases, and human medical and social conditions are typical responses to this need. As communications within a country develop and as regional economic disparities become evident to the population, the need arises to gather spatial data on a regional basis. The development of complex urban societies with their inherent strains brings a desire for control of spatial interactions, and thus a need for detailed spatial data.

The higher levels of economic development are characterized by intense competition between potential land uses and this creates the need for detailed spatial data on which to base decisions. Wider concerns of environmental degradation have initiated the gathering of data on conditions of wildlife habitat, recreational potential, and, more recently, the effects of man-produced pollutants.

It is implicit in the phrase "regional economic disparities" that countries, particularly large ones, do not develop homogeneously. Within any one nation there may be several distinct levels of economic development, and hence different perceived needs for spatial data in various parts of the country.

As well as variations in the levels of economic development, there are variations in types of political and social activity. Countries with a long history of self-government may have sets of data entirely different from those of newly independent countries. The latter may have data determined by widely different foreign policies of the previous colonial administrations. At the other extreme, some sparsely inhabited areas may be surprisingly well

surveyed simply because the political exigencies required evidence of territorial possession. Differences in political structure result in differences in data gathering. Strong central government necessitates good communications and efficient data gathering. Napoleon's innovative data gathering (and administrative) structure of French "Départements" still underlies collection of data in France. The politically oriented official data gatherer for each city block face during the period 1935-45 in Germany was an extreme example of collection of spatially precise demographic statistics, implemented by an authoritarian regime.

Military exigencies have had a strong influence on the perceived needs for spatial data. The supply of terrain data within any nation frequently reflects its involvement (either deliberate or involuntary) in warfare. It may well be, for instance, that Vietnam currently has the most extensive and up-to-date supplies of terrain data of any country in the world at a similar level of economic development. The influence of the military in data gathering is not necessarily bad. Perhaps the classic case is the Ordnance Survey in Great Britain, which has furnished that country (and others) with exemplary topographic maps. The first modern present-land-use survey of Great Britain was an indirect result of the stress of war. Similarly, the widespread planimetric mapping program of northern Canada was fundamentally dependent on the need for wartime aerial navigation. The military community has consistently nurtured the development of spatial data gathering techniques, including the earliest uses of aerial photography, photogrammetry, infrared photography, microwave triangulation, satellite sensor platforms, and laser distance measurement, all of which are routinely used in non-military spatial data gathering today.

Clearly, the perceived needs of government are complex and reflect the history of the country, its interaction with its neighbours, and its present mixture of levels of economic development.

Gathering constraints. Gathering constraints are conditions that affect the cost, length of time, or practicality of gathering spatial data in any country.

Physical environmental constraints are perhaps the most easily recognized, though their actual effects on the spatial data gathering process and the subsequent availability of various types of data are not so widely understood.

The size of a country does not prohibit detailed local spatial data gathering, but it does dramatically affect the cost and time of overall reconnaissance survey, which is a form of high-risk data prospecting. Nations with a large areal extent frequently spend a high percentage of their available data gathering budget on this level of survey. The extent and age of the different types of data available may thus vary with the size of the nation.

Climatic extremes of heat, cold, dryness, and wetness are well-understood physical constraints. Some of the side-effects of these extremes also act as important constraints. The subarctic and arctic areas are not only cold but dark for long periods of the year. As one approaches the magnetic poles, traditional forms of navigation are unreliable and radio communications are sometimes disrupted by electric storms, which are rare in lower latitudes. Humid tropical conditions not only restrict ground survey but the associated cloud cover inhibits aerial survey. The only widespread aerial view of Panama, for instance, has come only since cloud-penetrating p-band radar sensors have been employed.

Surface trafficability is clearly affected by constraints such as mountains, ice fields, sand deserts, muskeg, swamp, and various forms of dense vegetation cover. The author was once leader of a team of photo-interpreters who carried out a photo-geological study of 50,000 square miles of northern Ontario in Canada. The primary purpose of the work was simply to identify the rare rock outcrops in the muskeg so that helicopter-borne geologists could examine actual bedrock. Paradoxically, the same helicopters could not be used in the outcrop-abundant high mountain areas in British Columbia, because the air at high altitude was too thin for efficient use of rotary-wing aircraft. Dense vegetation cover not only hinders surface trafficability but can completely obscure aerial observation of terrain. Not a few contour maps of terrain height represent the surface of the vegetation canopy rather than the surface of the ground 0 to 300 feet below. Data gathering in large, sparsely inhabited or uninhabited areas involves a high cost for logistical support of the survey personnel. The effect of this is rarely fully appreciated. At a time when petrol was readily available at 50¢ per gallon in inhabited parts of Canada, for example, this fundamental commodity for transportation, warmth, and

cooking or fuel cost \$3.00 per gallon in the Canadian arctic. The cost of aerial survey for this item alone was thus six or more times higher per square mile in the uninhabited area than in the inhabited area to the south.

All of the above types of physical constraint contribute to the variability in type and supply of spatial data, and also affect the errors of observation and measurement in the data actually gathered. Similarly, there is a distinct difference in reliability between measurements of static and nomadic human populations, and between animal populations restricted to one feeding area and migratory ones.

Less obvious are the political and social constraints of data gathering within a country. A distinction is made between the constraints that affect the gathering of data and those that affect their use and availability after they have been gathered. The latter will be considered separately.

Within nations, particularly those with federal forms of government, there may be several clearly established jurisdictions, each with responsibilities for certain types of data collection. Within Canada, for example, each province is constitutionally responsible for data relating to its own natural resources. Where different political priorities exist within the various jurisdictions, different levels of spatial data availability can and do result. Different classifications may be adopted. Different units of measure, times of survey, levels of detail, and types of data gathering may result in a patchwork quilt of data availability within one nation. One of the major achievements of the Canada Land Inventory between 1962 and 1969, for example, was the performance of separate negotiations with each province that led to uniform classifications of present land use, agricultural capability, forest capability, wildlife suitability, and recreational potential of the land across Canada. If negotiations had been conducted by different personnel, in a different political climate, or at a different time in history, their outcome might have been less fortunate and the availability of spatial data in Canada would be measurably different.

Social constraints on data gathering also vary widely. The reliability of data collected by the U. S. Bureau of Census from black ghettos in the United States suffers in the same way as the census data from Corsica



gathered by the Government of France (INSEE). The hostility of a population to authority can be a severe constraint on data gathering. In contrast is the 400-year-old tradition of the Central Population Register in Sweden, where data collected on population attributes are not only detailed and generally reliable, but are freely available to the public. Similarly, the high rate of response to the postal method of census used in Canada would be totally impossible in a country with a high percentage of illiteracy.

The above examples do not exhaust the list of constraints that affect the gathering of spatial data. They are, however, sufficient to indicate that each country is unique. Each country has its own mixture of constraints that have to be overcome if specific types of data are needed, and that result in substantial variations in the type and character of data supply.

Gathering capability. The capability of a country to gather spatial data depends on the gathering techniques and the amount of funds available at the time when the need for the data is perceived. The rapid development of spatial data gathering techniques during the past century, and particularly in the past 50 years, will be examined very briefly below. Because of this recent development, the overall supplies<sup>1</sup> of spatial data available in various countries generally reflect their diverse needs in the past century (and hence their recent economic and social state) rather than earlier in their history. Diversity also stems from the fact that the application of different data gathering techniques results in different data products. For example, a geological map resulting from ground survey alone will probably be different from one produced by air-photo-interpretation, which will be different from one utilizing both techniques, which will be different again from one incorporating geophysical measurements of bedrock attributes. In short, different data gathering techniques provide products that have their own characteristics and limitations, and these affect the supply and nature of the available data in ways that may or may not be apparent to the user.

(a) Spatial data gathering techniques. Spatial data gathering relies upon the technologies of data sensing, transportation and communication, and

1. This applies to spatial data with a slow rate of decay, such as contour maps of terrain, geological maps, soil maps, etc., as well as to more volatile data.

data recording. Data gathering methods are the application of such technology.

The first of these factors, data sensing, can be defined as the ability to make an observation, and, by extension, the ability to calibrate the observation (provide a measure). The devices employed can be human<sup>1</sup> or instrumental. The relationship between them is illustrated in Fig. 2.4.

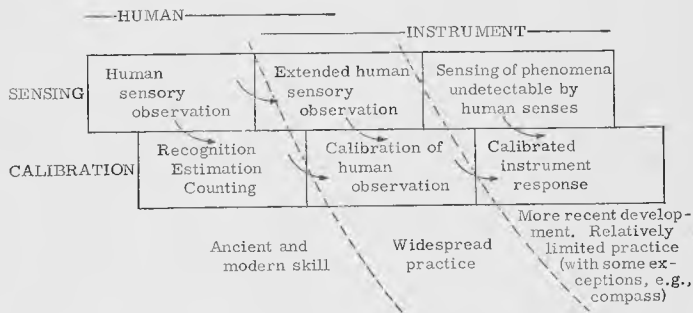


Fig. 2.4. Relationship between human and instrument capabilities for sensing and calibrating spatial data.

It is not necessary to catalogue the different capabilities, but several comments can be made on their range. Although the physical constraints on data gathering have remained more or less constant, the supply of the most common sensing device, man himself, has increased exponentially. It is a truism to say that more humans have lived, become literate, and participated in data gathering in the past 100 years than the total so occupied since the beginning of time. It is clear, however, that the growth of mankind not only generates demands for data, but also creates the devices for gathering them.

Instrumentation has extended the range of possible sensing, by extending human sensitivity through devices such as the telescope, microscope, and densitometer, which are sensitive to light; the thermometer, which senses

1. It is necessary to differentiate between human sensing capability and human data processing capability at this point. The latter quite powerful capability adds great utility to the relatively limited former one. The characteristics of human and machine-aided capabilities to handle spatial data will be examined in the following chapters of the thesis.

heat; the anemometer, which measures flow; the gravimeter, which responds to gravity; the echo-sounder, which uses and detects sound; and pollution detection devices which respond to chemical odours. Instruments also extend human sensitivity by sensing phenomena that man cannot detect, specifically energy wave forms such as magnetic fields, X-rays, gamma rays, radio waves, infrared and ultraviolet radiation, and chemical and physical properties such as mineral content and electrical resistivity.

Although all these types of sensing are used to gather data concerning attributes of the environment, human sensory skills (coupled with the instrument-aided extensions of sensitivity and calibration) predominate. The overall supply of spatial data now available has relied on the sensing ability of trained humans. The devices that sense phenomena undetectable by humans are not in such widespread use (with the possible exception of the compass); high levels of technical expertise are required to build and use such devices as magnetometers, geiger-counters, infrared-radiation sensors, and radar scanners, and similar high levels of expertise are needed to understand their results. They are, however, being rapidly developed and are capable of producing large volumes of data for subsequent interpretation.

Data may be gathered by either "contact" or "remote" sensing methods. Contact sensing occurs when the sensing device is in close proximity to the item being observed (as in interviews, ground surveys, metered traffic counts, surface meteorological recordings, or hydrological surveys). Sampling of soil, water, rock, or similar substances falls into this category because the samples are collected by a contact process. Remote sensing, on the other hand, relies on a sensing device far from the item being observed (as in airborne and satellite sensing of the earth, use of lidars and radars to measure atmospheric properties, echo-sounding of the sea floor, seismic measurement of subsurface materials, or spectral sensing of objects in space). Obviously the terms "close proximity" and "remote" are subjective, and represent points on a continuum of distance between the item being observed and the sensing device. The categories are thus not rigorous and only represent a convenient way to think of

sensing techniques. In similar general terms, whether the sensing device is a contact or remote one, the further step of recording the data may be done in close proximity to the sensing device or remotely, by means of some form of communication channel (as is the case with meteorological sounding balloons, automatic weather stations, or subsurface thermistors). These broad categories of techniques are summarized in the diagram (Fig. 2.5).

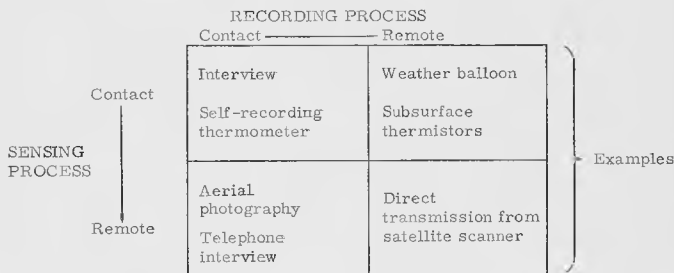


Fig. 2.5. Methods of spatial data sensing and recording.

This brief consideration of data gathering methods implies that the second area of technology, transportation and communication, is involved in the overall ability to gather data. Spatial data, by definition, are gathered from various locations in space. If contact sensing is employed, the sensing device must be transported so that it is in close proximity to the item or condition being observed. If the data are then remotely recorded, a communication channel is required. If remote sensing is employed, the sensing device must be positioned so that the item or condition of concern can be observed. If observations beyond the range of the remote sensing device are needed, the device must be transported. Again, if data recording is remote from the sensing device, a communication channel is required. The availability of spatial data thus rests on transportation and communication technologies. These capabilities have expanded enormously in the past century. The years since 1870 have seen the development of the automobile, dirigible, aircraft, telephone, radio,

television, helicopter, hovercraft, rocket, earth satellite, and early spaceship. Technology seems to have followed a pattern of accelerating growth. The central theme of this thesis is the handling of spatial data, but certainly the advances in transportation and communication technologies are major influences on the volume of spatial data gathered. We may be acquiring the ability to gather more data than we can handle; this point will be discussed in more detail in subsequent chapters.

The third technology with a direct influence on the ability to gather data is that of data recording. In the earlier part of the chapter the effect of the dimensions of the recording medium on the data format was discussed. The other essential element is the ease of imprinting a unique record<sup>1</sup> on a storage medium and the ease of its subsequent retrieval.

If a graphic or a symbolic language is used to generate a description that is recorded so that it can be read visually, a "graphic record" is produced. If either language is used to generate a description that is recorded on a storage medium from which it can be retrieved by other human senses or by machine methods, it is termed a "non-graphic record". It is obvious that data gathering is influenced not only by the ability to sense data, but also by the supply of devices, human and mechanical, that place the original record in a storage medium.<sup>2</sup> Less obvious is the fact that this ability is dependent on the availability of a suitable storage medium.

Compared with the present-day supply of data, few records were generated before the first part of the 18th century. Despite the advent of the printing press in 1588, paper was not cheaply produced until approximately 1750.<sup>3</sup> The widespread availability of paper, however, at a time of growing literacy, growing need for data, and growing ability to observe, strongly influenced the format and volume of the subsequent (and present) supply of

1. The process of placing a unique record on a storage medium is here differentiated from the process of reproducing copies of that record. The first affects the ease with which data are gathered, and hence which data exist. The second affects the dissemination of the data. These are considered to be primary and secondary influences in the total "availability" of data. 2. The important corollary to this, the ability to read either graphic or non-graphic records and to convert one to the other, is fundamental to spatial data handling and will be discussed in detail in the following chapters. 3. The first paper-making machine was produced in 1798. The first paper-making machine in Canada was built in 1803. Wood, as opposed to cotton or rags, was first used in 1720, was first used in a mill in 1800, and had generally supplanted other materials by 1880.

data. The later production of plastic film and the coating of metal, glass, paper, and plastic film with sensitive layers added greatly to the ease with which records could be generated. The period 1830-1890 saw the development of the light-sensitive coating and the birth of photography as it is known today. The impact of this development on the supply of available data is difficult to overestimate. Practicable photography coincided with the early development of the aircraft. Airborne cameras have been used routinely for over 50 years for remote sensing and recording of the earth's surface. The first Canadian aerial photography (for forest survey) was in 1921-24. Since that time, the whole of Canada (over 3 1/2 million square miles) has been photographed at scales at or larger than 1:63,360. This implies that there are substantially more square inches of original photographic documents available in Canada than there are of topographic maps at all scales. Even though sophisticated interpretation is needed to retrieve the inherent data from the photographic record (usually a non-coded image), the recording medium (in concert with sensing devices, transportation techniques, gathering constraints, and perceived needs) has had a dramatic effect on the spatial data available in Canada. Equally potent but still latent is the ability to store information on media with magnetically sensitive coatings. "Magnetic" wire, steel tape, and plastic tape became available between 1889 and 1935. These storage media, and data coded as patterns of holes punched in paper cards and tapes, form the basic repository of machine-readable, non-graphic records of spatial data available today. The ability of machines to read and reproduce such records quickly, and subsequently to examine and manipulate their values, is a central concern of this thesis and will be examined in the following chapters. It is sufficient to say at this point that data recording technology has a substantial impact on the ability to gather spatial data and on the supply of data.

(b) Available resources. The concept of perceived need expressed above included the decision of a country to allocate resources to data gathering; to pay for the data. To some extent, the internal availability of resources in a country has an effect on the supply of data, but the relationship between available funds and available spatial data is far from clear and certainly not linear.

In general, a developing country, "poor" in terms of per capita income, will not have the ability<sup>1</sup> to gather as much data as a "rich" country. In the former, strong competition exists for its scarce resources of trained manpower and technical expertise; social constraints such as illiteracy may be significant; and administrative networks may be rudimentary and unable to gather reliable data. This latter point is worth amplification. The governmental sources of spatial data can be generally subdivided into "administrative" and "survey" categories. The former produce data as by-products of the administrative activities of government (tax returns, building permits, vehicle registrations, health records, crop yields, and so forth). The latter are the result of specific decisions to gather the data (census, soil survey, geological survey, meteorological survey, and others). In a developing country the administrative structure may generate little administrative data, whereas the greatest volume of spatial data available in an industrially advanced nation may come from such sources. Survey, on the other hand, requires less well established government structures and can take more rapid advantage of advances in data gathering technology. Advanced data gathering techniques, however, have been a product of the developed countries, and they are usually first applied to the benefit of such countries. Nevertheless, the technology can be readily exported. A very common form of international aid is the provision of survey expertise; the author has taken part in several such missions. The very process, however, underlines the difference between administrative and survey data.

The type of records generated by administrative data can be readily envisaged; their generation is directly related to the sophistication of the administrative structure in a country. Survey data can employ a wide variety of gathering techniques and, after taking the perceived need and gathering constraints into account, the availability of the products of surveys in any one country depends as much on political decisions, international alignments, and incidence of warfare as it does on the internal availability of resources.

#### Resulting Data Characteristics

It is clear from the above discussion that spatial data are gathered for a

1. And probably not the same need.

variety of reasons, by various methods, and in the face of various difficulties. The approach used to gather a specific type of data will obviously affect the nature of the records produced and their reliability. The set of choices made in a country determines the characteristics of the overall supply of spatial data to be handled, and hence the types of handling technique that would be useful.

The use of data is affected by the characteristics that constrain their validity and those that influence their physical manipulation. The range of these characteristics is considered below.

Validity constraints. All the components (attributes, spatial and temporal characteristics) of a data element can be considered to be at some level on an imaginary scale of validity (truth). They may not necessarily be at the same level, and their actual reliability may or may not be apparent.

The first aspect of validity is the level of error that may have been introduced by the process of observation and measurement. No human or instrumental sensing device is perfect; the errors produced may not be random, and environmental conditions may differentially affect the result. The sensing process itself may affect the subject; an inquiry by a government agent to a farmer concerning his crop damage, for example, may yield data entirely different from those obtained by airborne sensing of the same information. The range of such error for any one observation can obviously be large. Careful design and screening of data gathering processes can, however, reduce the overall error of a data set, and with adequate procedures the degree of error in a data set can be estimated.

The second aspect of validity concerns the nature of the entity or condition being observed, and the use to which its description will be put. It was mentioned above that only when objects or events are countable by the techniques of observation and measurement used can a complete observation be made. In all other cases, and in many instances where the subjects are countable, a sample is taken. An examination of the theory and practice of sampling is outside the scope of this discussion. However, the way in which the sample is taken and, in particular, the frequency (density in spatial terms) of the sample, clearly have impact on the completeness



with which an entity is described. The resulting level of description, coupled with the level of error, may or may not be significant to the decisions to be made from the data; the data may or may not be valid in terms of a specific use.

Several broad categories of level of description can be recognized. They are essentially based on differences in sampling frequency with respect to the nature of the entities being observed, as outlined below.

Exploratory data gathering results in general wide-area inventories based on broad classifications. Using natural landscape features as an example, major topographic-climatic zones are identified and major relief units are delimited. The results are valid for an initial assessment of the potential for commercial development.

Reconnaissance data gathering typically results in data that can be used to confirm the existence of possible commercial opportunities and to identify areas that warrant more detailed study. Using the natural landscape example again, great soil groups, major land forms, and vegetation communities are differentiated.

Intensive data gathering provides information that underlies preliminary capability studies, cost benefit analysis, and so forth. Soil series, detailed land forms, and vegetation associations are identified in the natural landscape.

Detailed data gathering provides information at the maximum level of detail needed to establish reliable cost-benefit ratios or to make input into management decisions. In the natural environment, phases of soil series, land form elements, slope units, vegetation species distribution, tree counts, and so forth, are identified.

The level of description is clearly related to the intensity of data gathering. The above categories are essentially simplistic subdivisions of a continuum. The prevalence of data sets that provide specific levels of description in any country obviously depends on the data gathering constraints, and the need and techniques used to overcome those constraints. Individual data gathering techniques have a range of applicability to certain levels of

intensity of data gathering, depending on their inherent sensing resolution and the effort needed to employ them.

The difficulty arises when data gathered at a low level of sampling frequency or inherently high level of error, or both, are used for decisions that require more detailed or accurate data. Further difficulties arise when data with various sampling frequencies and sampling designs are compared or combined. These are elementary concerns, but have direct impact on the utility of subsequent techniques of data storage and manipulation.

All data have a rate of decay of validity. This may be specified to some degree by the temporal identifier attached to the data element or data set. It may not be specified, but may be understood to be a generally constant rate of decay (as for a population census). Other data, however, such as political opinion, may be subject to unpredictable changes in the rate of decay and there may be no appropriate measure of this "volatility".

While it would be valuable, it is not within the scope of this chapter to examine fully the theory of data classification. In general terms, "primary" and "secondary" data can be identified. Primary data are direct records (either contact or remote) of the sensing device response. Secondary records are generated by processes such as generalization, interpolation, or interpretation of primary data, usually within some specified rules of classification. The difficulty arises when the processes are mixed to produce one secondary record, as they frequently are. A soil map, for example, can show soil types that, within one sheet, may have been delimited by generalization of numerous ground observations, interpolation between sparse ground observations, or interpretation of aerial photographic images, or any combination of these techniques. On the other hand, such classification may reduce the effect of error in individual data and generate secondary data that are adequate for decision making by the user. As a result of these processes, however, there may not be an adequate measure of the internal consistency and reliability of a secondary data set, and its validity for specific decisions may be effectively obscured. The problem is that secondary records are not infrequently the only "source" data available.

Physical handling characteristics. The choice of techniques and level of activity for data gathering in a country will determine the physical handling characteristics of the resulting primary data. The choice of language and recording medium for descriptions will determine the ease with which the resulting data may be read. The intensity of data gathering and the size of the country will determine the volume of data to be handled.

In the early part of this chapter (see "Spatial data format") it was stated that languages for describing spatial data are expressed either in a symbolic form (alphanumeric plus special symbols) or in a graphic form. Within the latter it is possible to differentiate further between diagrams and pictures, diagrams being graphic expressions of a known spatial language, and pictures the uncoded records of images. (It is an open question whether the latter is an expression in an undefined graphic language.) For practical purposes, the media currently used for recording spatial data are amenable to the production of either graphic or non-graphic records. The first are visible marks on documents or are photographs (on paper, plastic film, or, less commonly, glass or metal). The second are machine-readable marks, such as punched holes in paper cards or tapes, or magnetic signals on the coating of tapes, disks, drums, and so forth.

The various types of data sensing tend to produce specific data formats, but there are exceptions within each category. Human sensing and extended human sensing, for example, generally produce primary data in a symbolic format. Many of these are transformed to secondary data in a graphic format. An exception to the general rule, however, is the surveyor who plots thematic boundaries on maps as he walks the ground, or who draws sketches of rock contacts and so forth, thus producing data directly in graphic format. Self-recording instruments usually produce data in graphic format as diagrams (graphs or tracings), but some now produce digital data (in symbolic format). Sensing devices that transmit information for remote recording usually send data in codes (a symbolic format), which may be subsequently transformed into an analog or digital form (graphic-format diagrams or symbolic format). Photographic sensing produces data in picture format. The mixture of data

formats produced obviously depends on the data gathering techniques employed. However, data in a graphic format can be stored only in graphic records, whereas data in a symbolic format can be stored in either graphic or non-graphic records. If data in a graphic format are to be stored on non-graphic records, they must (in 1974) first be converted to a symbolic format. These possibilities are illustrated in the diagram (Fig. 2.6).

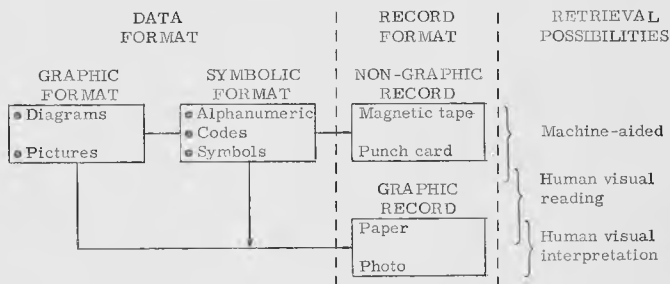


Fig. 2.6. Relationships between data format, record format, and retrieval possibilities.

The record format in turn determines the facility with which the data can be retrieved for use. Graphic records require human visual retrieval, and photographic records may require a sophisticated process of interpretation involving human pattern recognition before data are identified. Non-graphic records allow the use of machines to read the data rapidly. These capabilities overlap somewhat: a limited number of machines (optical character-recognition devices) can read symbols on graphic records, and equally, a human can interpret a pattern of holes in a punch card or tape and slowly determine the data content. In terms of current practice, however, these are unimportant exceptions.

In general, human retrieval of data from graphic records is a slower process than machine retrieval of data from non-graphic records. The effect of this on the use of data and the relative capability of human and machine-aided techniques for subsequent data manipulation will be the subject of the following chapters.

The volume of data available is probably the most important influence on the way the data are subsequently handled. In countries with major data gathering constraints and a low level of data gathering ability, the amount of spatial data generated is likely to be well within the existing capability for handling by humans or machines. However, although the constraints on data gathering are slow to change, the ability to gather data is increasing rapidly, and so is the volume of spatial data. The volume of data generated in graphic, as opposed to non-graphic, records, becomes particularly significant because it requires human retrieval, which is relatively slow. The need to supplement human data retrieval and manipulation with machine-aided processes clearly depends on the volume of data generated in the country concerned.

Use constraints. The need for spatial data may have been perceived, the constraints on gathering may have been overcome by various techniques, and the required data may already exist in some format. For several reasons, however, there may be constraints on their actual use.

The constraint on data use imposed by their validity was mentioned above. The level of error, sampling design, or age of data that apparently "exist" may make them unsuitable for a particular use.

Other constraints on use can be recognized. Various types of spatial data may be required for a specific purpose; they may all exist, but may be dispersed in numerous departments, at various levels of government, in various parts of the country. The effort involved in assembling the required data sets may exceed available resources, and, in practical terms, the data cannot be used. High data volume combined with low reading capability may have the same effect. If a decision has to be made, it may not be possible for assembled data to be read, with available resources, in time for their information content to be brought to bear upon the decision. Again, in practical terms, they are not available.

A more pervasive influence on the availability of data is the constraint of "confidentiality". In many countries, government departments gather data that thereafter are legally confidential. There are several dimensions to confidentiality. Legal confidentiality of certain types of data is prescribed

to protect the privacy of individuals (as in the case of census or taxation data), and such data may be made available only in summary form. Other confidentiality constraints are imposed merely because the data gathering agencies wish to control the data, because the collection or storage have been poor, or because the data may be critical of the institution's policies. There are many cases of apparently innocuous data being suppressed for no clearly defined reason. The military, in particular, have a strong influence on data availability in some countries. Their attitude stems from a desire to prevent data from reaching the "enemy" and to avoid revealing their own data gathering capability. In North America, for example, it is speculated that at least 50 earth satellites occupying various mobile and stationary orbits are monitoring conditions on the earth's surface. Some of them are reputed to operate sensors with high levels of resolution. The only satellites producing widely available spatial data are the TIROS and ERTS satellites, and they carry sensors with only low-level resolution.

One can argue that it was always thus and that these constraints are a function of the nature of mankind. It must, however, be pointed out that such constraints are a further variable that influences the supply and availability of data in any particular country.

#### Resulting Spatial Data Supply and Availability

The variables that determine spatial data availability described above are clearly interdependent. The availability of a particular data set in one country is the result of a series of choices made with respect to each of the variables. The series of choices can be thought of as a path traced through the various constraints and capabilities.

A geological survey in Canada, for example, would represent the perceived need for data by the central government of a country with a relatively high level of economic development. The size of the country forms a constraint; also, the climate is generally cold and overall trafficability is bad. There would be little political or social objection to the survey. A balanced combination of human and extended human sensing devices could be employed, transportation of these could be aided by aircraft, helicopter, and

automobile, and both remote and contact sensing would be employed, with contact recording in each case. The recording media would include both paper and photographs, and the survey would be undertaken by a survey agency using trained residents. The data characteristics would result from a substantial amount of reconnaissance data gathering, a fairly low level of error in primary observation, and a medium level of error in secondary records due to the relatively low level of intensity of the data gathering. The resulting data would include large numbers of annotated air photographs, plus documented field notes as primary data, and a relatively few small-scale maps as secondary "source" data. All data would be available for general use on request. The maps would be available as a printed map series.

As a contrast, let us consider the influences on a recent geological survey in a country such as South Vietnam. The country has a low level of economic development, and the survey would be undertaken primarily for military purposes by a central government. It is a medium- to-small country with a hot climate, extensive vegetation cover, and poor overall trafficability. Political constraints may be low, but social constraints on a survey may be high. The technique used would be almost exclusively photographic, airborne, remote sensing with contact recording resulting in photographic records. The survey would be undertaken with technology derived from outside the country and carried out by nonresident technicians with little experience of the ground conditions in the country. The resulting data would probably include a medium amount provided by reconnaissance survey, and some at a higher level of data gathering intensity from sites of strategic importance. A medium-to-high level of error could be expected in primary and secondary records. The resulting data would include large numbers of annotated air photographs, few or no supporting field notes, and a medium-to-high number of medium-scale maps as secondary "source" data. There might be severe constraints on the availability of the data for general use.

A third example might be a geological survey of a country similar to Great Britain. The country is assumed to have had a perceived need for geological survey for commercial purposes during the industrial revolution, and

it is under a strong central government, with no direct military influence. The country is small, has a temperate climate, presents good overall trafficability, and exhibits no significant political or social objections to the survey. The early initial perceived need for the survey, coupled with the low level of constraints, would allow human contact sensing and recording, by trained residents employed by a survey agency. The resulting data would represent a high level of data gathering intensity, a low level of error, and the production of a medium quantity of large-scale paper maps as "source" documents. Field documentation would exist, but would not be available due to gathering-agency constraints.

The three examples illustrate typical differences in the data characteristics, record characteristics, and availability of data that exist for one type of data in different countries. This variability occurs not only between countries, but between parts of countries, and between data types within one country.

#### SUMMARY

This chapter has briefly examined the characteristics of location-specific data. The concepts that underlie spatial description were considered, and the influence of the storage media on spatial data format was shown. The main types of location identifier currently used were identified, the factors that influence the overall supply and availability of location-specific data were examined, and the range of variation that exists was illustrated.

Once a supply of spatial data has been obtained, the need exists to store, read, and analyse them if their information content is to be applied to decisions. The next part of this thesis will examine the state of development of spatial data handling techniques. The usefulness of such techniques depends to a considerable extent on the supply of data that must be handled, and it has been shown that this may vary substantially between countries.



### CHAPTER 3 - MANUAL TECHNIQUES FOR HANDLING MAPPED AND OTHER LOCATION-SPECIFIC DATA

For the purposes of this discussion, manual techniques for handling mapped and other location-specific data are taken to be those accomplished by human skill, aided by drafting and photographic equipment but not by electronic data processing equipment. In general, they form the basic methodology of long-established cartographic practice. They are still used to carry out most spatial data handling tasks. Also, viewed in the total context of data handling rather than only as "traditional cartographic practice", the various manual methods form a series of benchmarks of traditional capability against which machine methods must be judged. Descriptions of manual techniques in this chapter will focus on their timing and efficiency.

Viewed in the context of data handling, a map is a location-specific data "file" in which earth data are gathered, ordered, and recorded in a way that preserves the measures of their spatial attributes and distribution. The file's structure presupposes some conceptual model of the space occupied by the globe, and suitable units for its measurement. The Euclidean model, with degrees of latitude and longitude as units of measurement, is the one widely accepted. Its appropriateness, utility, or validity will not be called into question in this chapter.

The classic task of geodetic and topographic surveying is to record in such a file the configuration of the earth's surface and a very limited set of its features. The location of a series of identifiable points on the ground is established by multiple measurements between the points, and between them and extraterrestrial bodies. Plane graphic representations (maps) of the curved surface of the earth are produced by mathematical transformations (map projections). Established points are "filed" on the maps according to their measured relative positions. Elements of thematic data are located by observing or measuring their relationship with easily identified features already stored in the spatial file, and are recorded by inserting them into the appropriate place in the file, that is, by plotting the observations on a map.

The use of aerial photographs as the source of both topographical information (photogrammetry) and thematic data (photo-interpretation) follows the same general procedure as for ground survey. The main difference is that the relationships between objects or the relationships between thematic data and topographic objects are observed and measured on photographs, and only a few observations on the ground are necessary.

Before 1960, the only form of storage for spatial data was graphic (as maps), and its precision varied. Data compaction was attempted by use of symbols, and data separation and identification were frequently aided by colours. Retrieval of data from the file was mainly manual. These manual retrieval procedures have not changed substantially in the past 30 years.

The starting point of manual procedures for handling graphically stored data is thus the map itself. The efficiency of the map as a source of data from which useful information has to be extracted, and the techniques for extracting it, need to be examined. Such spatial data handling techniques fall into three general categories. The first is "image manipulation", which changes the area of the graphic material on which data are to be displayed. This category includes changes of scale, correction of distortions in the original material, and changes of map projection. The second involves the "grouping" of data or sets of data and the production of a new or amended record prior to data extraction. It includes such procedures as centroid allocation, merging and dissolving of data, and generalization. Both image manipulation and grouping are broad categories of data handling techniques that affect the retrievability of data from the graphic file.

Although there is inevitable overlapping between the two groups, the former is primarily concerned with changes in the graphic image of the data and the latter is concerned with changes in the substantive values of the data elements. The third general category is the process of "data extraction" itself, and includes visual inspection and estimation, selected extraction, and measurement.

The storage medium affects all subsequent processes of data handling. It may vary from paper of various qualities and conditions to more stable

materials such as paper bonded to metal or plastic films (opaque, translucent, or clear), emulsions on plastic or glass, or direct engravings on metal. Each of these media has somewhat different physical characteristics, and may be used at various stages of map production. As an end product, however, there are two predominant types of graphic data storage: the map that is colour-printed on paper, and the map drawn in a single tone on plastic film. These two types of graphic storage encompass the overwhelming amount of spatial data in graphic form.

#### IMAGE MANIPULATION

Image manipulation is the controlled stretching or shrinking of a given image. It is achieved by various means and used to affect the retrievability of the graphic information. Simple change of scale is a result of equal and orthogonal stretching or shrinking. After reduction of the image, information retrieval is affected by factors of visual acuity, and after enlargement by the factor of visual scanning (reduction of synopsis). Rectilinear distortions<sup>1</sup> in a given image can be eliminated by linear stretching or shrinking along one or more axes (not necessarily equal or orthogonal). Nonlinear distortion elimination is a nonlinear version of the latter process. Map projection change is essentially a controlled form of linear or, more often, nonlinear stretching or shrinking where the properties of the distortion in the resulting display must be known. The subsequent retrievability of information may be affected by both of the factors relating to scale change noted above, and also by the change in distribution of substantive information.

Linear stretching or shrinking of an image can be achieved in several ways. The simplest but most laborious is the point-by-point transfer of data by hand and eye from the source material to the new enlarged or reduced format. More common is the use of a "pantograph", a device that mechanically links a drawing device to a cursor with which the original image is traced. The linkage is adjustable so that the image being drawn can be larger or smaller than the original. Optical methods involve projection of the source material on to a suitably aligned plane surface, where the new image is recorded by tracing or by using photosensitive material. The use

1. Elimination of rectilinear distortion in this sense does not mean conversion to the all-curved, spherical surface but refers only to distortion in comparison with a standard that has not undergone dimensional change.

of an epidiascope or simple slide projector to shine images on a flat surface falls into this category. The inherent distortions of such simple optical systems can be removed by the use of high-precision copy cameras, stable base materials, vacuum frames, and so forth. The enlargement or reduction of letters, numbers, and symbols does introduce problems of aesthetics and legibility, and if these become severe, the textual information may have to be removed and reintroduced after alteration. Colours can be transferred from the originals by use of colour photography or colour separation and a further printing process. Nonlinear stretching or shrinking of an image cannot, however, be achieved with such simple optical or mechanical systems; it requires quite expensive optical processes in which individual lenses are produced for specific nonlinear corrections, or else the simpler but slower process of point-by-point transfer.

#### Scale Change

The current manual processes of scale change thus involve either relatively simple procedures with inherent inaccuracies (in terms of human error, in optical systems, and in tracing processes that mechanically transmit or fix the image), or they employ relatively sophisticated, high-precision optical equipment, experienced technicians, and not inexpensive photosensitive materials and chemicals. To change the scale of a 20-in. x 30-in. map sheet manually would normally take between 1 and 10 hours of experienced labour, depending on the data content of the map sheet concerned.

#### Distortion Elimination

Although distortion elimination is usually performed when data are inserted into a spatial file rather than while they are in their stored form, techniques involving the latter illustrate the difficulties of manual processes. Non-stable material such as paper may occasionally be used for storage, and it may stretch or shrink because of changes in temperature or humidity. If such a distorted data set has to be overlaid on a nondistorted one, the distortions must be eliminated. Only linear distortion can be effectively corrected by economic manual processes. A transparency of the source material has to be produced at a scale that can be accepted by a projector, and this process itself may introduce inaccuracies, particularly if the

transparency has to be small. Production of large transparencies reduces the error, but requires copy camera equipment. The image can be made to fit control points drawn on the image plane, by a process similar to that of rectifying photographs. Because of the nature of the process, and because it is difficult either to correct nonlinear distortions or to assume that the distortion of the source data is in fact linear, the process is rarely used. Minor distortions in graphic source data are usually ignored, or at best "allowed for" when mapped data are handled manually.

#### Nonlinear Distortion Elimination and Map Projection Change

These procedures do not lend themselves to optical or photographic processes, so the transfer of data from source material to a new format is achieved by eye and hand. Calculated reference points are used as a guide, followed by interpolation between data elements when sufficient numbers have been transferred; if great care is exercised the accuracy can be high. The time taken mainly depends on the number of data points transferred. A single 20-in. x 20-in. outline map of political boundaries and main physical features of western Europe, for example, would take more than one man-day. The effort required for the manual method means that it is generally undertaken by cartographic houses rather than by research geographers.

#### GROUPING

Grouping encompasses several techniques used to change the form of the data in a map prior to data extraction.<sup>1</sup> It includes the relatively simple concepts of merging data or sets of data and dissolving redundant data elements (for example, removing a line that separates two areas of the same kind, which may have resulted from either merging or matching edges). It also includes "modal change" and the "generate" process, where data elements are grouped or represented on the map by a newly generated data element or elements (e.g., centroid allocation, contouring, generalization).

All these techniques involve selective processes as well as purely mechanical reproduction; the latter may be harder to accomplish than the former.

1. Both "image manipulation" and "grouping" involve the creation of a new record, i.e., a new map, and, in that sense, could be termed "map compilation". The latter term, however, is frequently used to describe the assembly of spatial data to produce an original map. As this discussion concerns itself with techniques that are employed only after the original map has been produced, the term "map compilation" has not been employed in their description to avoid confusion.

### Merge and Dissolve

To merge is to combine contiguous areas that qualify for inclusion in a new category or set of categories, and to dissolve is to remove a common boundary between specified areas and to group them in a single area. Either the entire file can be traced (or photographically reproduced) with the appropriate elements omitted or masked, or the original file can be amended. Except when minor changes are made to an existing file, either process is rather laborious. The material of the source document, the number of changes, and the neatness required in the final document have a substantial effect on the time involved. Stable-base photosensitive materials have been developed on which data from translucent single-tone source material can be copied and from which data can be removed by electrically powered erasure. However, if the source material is colour-printed paper and colour washes must be changed in the merged areas, the document must be completely reprinted using corrected colour separation plates.

### Edge Match

The same difficulties have to be faced when the edges of adjacent maps are matched. To edge-match one sheet, eight other sheets must be examined visually and corrections may be required on four or more of them. The time taken varies with the number of mismatches to be corrected and number of references that must be made to data sources, but the level of effort required can be understood.

The brief discussion of manual techniques for handling spatial data so far has indicated the relatively laborious nature of certain of their operations. This particularly applies to the procedures involving the transfer of data from one medium to another, especially if it is not a linear process. On the other hand, the selective process undertaken by the human mind as an integral part of manual graphic data processing is often extremely efficient. This is illustrated in some of the techniques described below.

### Centroid Allocation

Centroid allocation is the determination of the centre of a data cluster according to a given algorithm. The algorithm can prescribe whether it be the centre of mass, a centre constrained within the boundary of the area,

or the centre of areas, lines, clusters of lines, or clusters of points. The eye, hand, and brain can perform this task easily, as it involves pattern recognition both to calculate the position of the centroid in the data cluster and to separate the data cluster from other elements in the spatial file. Calculation of the centroid, a rather complex process in terms of the decisions involved, is done almost instantaneously. Considerably more time, however, is needed actually to plot the centroid on the graphic medium. On a typical 20-in. x 30-in. map of present land use at a scale of 1:50,000, with 2,000 patches of homogeneous land use, each of which requires a centroid, the process takes approximately 2 hours for point allocation (3 hours if the points are sequentially numbered on the map as they are inserted), and a further 1 hour to check that none have been missed. The same efficiencies apply in the tasks of manual contouring and manual generalization.

#### Contouring

Contouring, the creation of isolines through a series of points with variable values, relies upon visual examination of the original data set and a decision-making process to direct the drawing of isolines through the points by hand. Conflicts of interpretation may or may not be resolved by reference to additional data. The manual approach has the advantage that it makes immediate use of the relevant knowledge of a draughtsman.

#### Generalization

In generalization, elements of spatial data, possibly with multivariable classifications, are combined into a single larger element with a single classification, usually based on the proportion of the original separate classifications contained in the new spatial data element. The process is essentially a sophisticated sequence of merging, dissolving, and generating calculated according to a set of rules. The technique is used when the complexity of data prohibits sensible display at the desired scale, and is of particular significance when the resulting graphic display is to be the only place where these data are stored. Generalization is part art and part science. The degree to which data can be crowded on a particular map sheet is frequently determined by mental evaluation of the ease with which the map sheet can be read, or the overall aesthetics of colour or line patterns. Here again, the mental processes of graphic selection are

extremely efficient. Once the decisions have been made, the new boundaries may be traced selectively from larger-scale sheets, the images reduced photographically, and the data elements relabelled; lines may be smoothed; or, for extreme generalization, the original data may be symbolized. The techniques used in these different approaches may vary, and can be mixed on the same sheet. All the physical operations are laborious and involve the creation of a new spatial data file. The time and effort taken necessarily depend on the area to be covered and, in particular, on the rules of the generalization.

#### DATA EXTRACTION

Data extraction is the dual process of selecting and retrieving data from the graphic storage file for subsequent decision-making purposes.

"Data selection" depends on the requirements of the particular decisions to be made, and is based on a priori knowledge of the contents of the file, or on the heuristic process of browsing through the file (interactive file search), or, more usually, on a combination of the two approaches. The mental processes of such identification and selection are essentially logical comparisons of the file elements with mental sets of data, or rules prescribed by a decision-making process. As the ability to browse through graphic storage files relates to the physical process of data extraction, it will be briefly examined in this section of the chapter.

"Data retrieval" encompasses the direct reading of data elements written in the spatial file and also the retrieval of data inherent in the structure of the file contents. This latter aspect of data retrieval incorporates measurement and comparison, which are controlled by the data selection process and which can be either mental (estimation) or physical. Either type, mental or physical, can be used to identify data elements, identify and retrieve data patterns, compare separate data elements or patterns, and superimpose data elements or patterns. Data retrieval is completed by moving the selected data elements or results of the measurement or comparison to another form of storage, either mental or physical. The mental operations involved in measurement and comparison are outside the scope of this thesis, but the physical processes of measurement and comparison



form a substantial part of the techniques for manually handling graphic data. The manual operations involved in these processes include the ability to read, search, move (transfer data from one record to another), measure, and compare. Some examples that illustrate the time and effort involved in these tasks are given below.

#### Data Selection

Reading. Reading is by no means simple and straightforward. Character recognition is a complex process that the human mind can do well, given adequate training, but that computers cannot (hence the need to change graphic data into special machine-readable codes as an essential first step in machine data handling). Furthermore, reading graphic files implies the ability to interpret the symbols and particularly the conventions used in the cartographic displays. It requires mental adaptation to the constraints of these graphic languages when estimation and measurement of the graphic file are involved. It is not unusual for the reading of a 20-in. x 30-in. topographic map sheet, undertaken to determine printing errors, to take 10 hours, and the time is correspondingly longer for sheets with high densities of data. Such reading for editing purposes does, of course, include the time used for mental comparisons with a given set of editorial rules, but it indicates the generally slow nature of this basic aspect of manual retrieval. The reading process itself is not free from errors. It is common practice to use more than one editor for map checking, not only because one person can introduce errors into the information as it is being read, but because one person can omit to read part of the file attentively, even though it has been visually scanned by that person.

The time involved in reading a particular graphic data set is a function of reading capability and file size. Although map sheets are thought of as a compact and symbolic form of data storage, the number of sheets required to contain even a limited amount of the spatial data already gathered from a particular area of the earth may be large. This can be illustrated by two examples. The five data sets (present land use, agricultural capability, forestry capability, recreation suitability, and wildlife habitat suitability) collected by the Canada Land Inventory from the agricultural and productive

forest areas of Canada (approximately 1 million square miles) will comprise 15,000 to 20,000 sheets when complete (14,000 have already been produced). Of the more than 500 electricity utility companies in the United States, a typical company has 10,000 to 15,000 map sheets to record the position of its power line network. Even the physical storage requirements of such data sets are not inconsiderable, and users frequently have a subset of graphically stored material for regular use and central (remote) library facilities for main storage. This measurably lengthens the access time to the total store and has some influence on the data readily available (hence most used) for decision making. Finally, the actual handling of the files causes eventual physical deterioration of the images and regular maintenance is required.

Search and identification. The time and effort involved in all types of "search" operations, both manual and computer-aided, are obviously influenced by the way in which the data elements are organized and indexed.

The storage of graphic data imposes several significant constraints on the way spatial data are filed and on the volume of data on any one record. Attempts to file the many different types of individual spatial data elements on separate sheets and link them by logical indexes rapidly create prohibitively large numbers of sheets. To minimize the number of sheets, sets of related, but different, data elements are frequently filed on one sheet; for example, a topographic map shows rivers, buildings, forests, contours of height, and so forth. Alternatively, various data elements may be combined and classified according to a set of rules, and only the resulting categories are shown. These procedures are useful if the user wishes to consider the resulting data sets in their entirety, but they can make it very difficult to search for subsets of the originally gathered data, and where reclassification has occurred may even prohibit their extraction.

The file structure used, however, may greatly facilitate the search process. In some instances, such as a topographic map, the file structure of the data elements conforms approximately to our perception of the various environmental elements, and their relationships to each other and to our behaviour patterns. This provides us with a series of mental indexes which establish

a hierarchy of values for data elements and keys to their spatial distribution.<sup>1</sup> Use of such established mental indexes can measurably reduce the amount of reading required in the search process.

Conventional physical indexing of graphic data storage is usually limited to broad groups of subject matter, cross-indexed by spatial attributes. Little can be accomplished by reference to such simple indexes alone, but they provide rapid identification of data sets for specific areas; subsequent access to the map sheets is direct (random) and the time used can be short. The storage medium is the same as the display medium, so data retrieval can begin instantly. Browsing is greatly aided by the random access to map sheets. It is easy to display several map sheets in rapid succession and to view them side by side. For small numbers of sheets this is an extremely efficient operation, but if many maps are displayed at once it may become difficult for the observer to recognize and retain the images they contain. This may be a function of low human "channel capacity".<sup>2</sup> The process is, however, entirely under the control of the observer and becomes cumbersome only if large files have to be scanned. When the simple indexes allow identification of limited search areas and the available data sets are conveniently grouped, the manual process is rapid, efficient, and inexpensive.

The end product of the search and identification process is selection of the data elements to be extracted from the file. In mental terms, this involves matching the pattern of the graphic elements to a given set of rules or to images based upon a set of graphic rules. Where the data element, for example a name, is written in the file, identification is a straightforward matching process. Where the data element, for example the distance between objects, is inherent in the file structure, the processes of retrieval by measurement or comparison (either mental or physical, or both) must be carried out to reveal data that match a selection requirement. These latter operations will be examined further below.

#### Data Retrieval

Transfer. The transfer of data elements from a graphic file to a second storage for use in decision making may be accomplished either mentally

1. Milgram 1972, Emerson 1969, Stea 1969, Trowbridge 1913. 2. Human "channel capacity" is the maximum rate at which bits of information can be transmitted to the brain via all human sensory channels.

or physically.

As the format of the mental storage is largely unknown, no sensible comment can be made about the mental process of transfer to that form of storage except that it is rapid if the data are simple, involves only labour costs, and has proved its ability. No measures of image retention capacity are available, but it is thought that at least 250,000 images are filed in storage as part of the normal human visual recognition process. The process incorporates random access to such images, a complex topological index, browsing capability, a sophisticated image-matching capability, and an individual image matrix of 1,000 by 1,000 points. The rate of human data transfer is considered to be in the order of 3 baud,<sup>1</sup> which would indicate a rather low "channel capacity" coupled with a high processing capability. No estimates are available for the modal division of such data between the various sensory channels. These figures cannot be substantiated without extensive further investigation, if at all in the foreseeable future, but they probably are minimums (particularly in the case of channel capacity). They indicate the formidable capacity of the human system of image storage and manipulation, which must be kept in mind when alternative machine-aided systems are considered.

The physical movement of selected graphic data to a second storage is an image transfer process accomplished by a variety of reading and writing operations. The operator may read the data element and write it on a separate record. If the image is other than symbolic, the error level of such visual transfer may be high; and, as an alternative, transfer may be made by contact tracing on to a second record, or by projecting the image on to the second record and tracing the light image. If the images are extremely complex and tracing would result in an unacceptable number of errors or expense of time, then a photographic transfer process may be used.

The advantage of manual tracing is that unwanted data can be mentally masked and only selected data need be transferred. Photographic data transfer may make use of physical masking, in which unwanted parts of either the original image or its photographic copy are covered. However,

1. Edwards 1964, Quastler 1956.

this task itself may be more laborious than tracing. A filter can be used in the photographic process to mask specific colours on coloured originals. However, physical transfer of data from a graphic storage file to a second store is usually accomplished by tracing.

Experimental evidence that reveals the degree of error and the effort required in the process has been obtained in tests of data transfer by tracing and scribing techniques performed by the Cartographic Section of the Canada Department of Agriculture in the Government of Canada. Different amounts of data were contained in the graphic storage file for various parts of the country. Accordingly, samples comprising one file containing data of average density and another containing data of low density were chosen for the tests. The first consisted of a soil map and a present land use map of part of Grenville County in Ontario, both maps at a scale of 1:126,720. Two tracings from the present land use map of Grenville County were made, and the tracing times for neat line boundaries and present land use area boundaries for the entire map sheet of 339 sq. in. were as follows:

a.	Experienced draughtsman	4 h 35 min
	Editing of tracing process	1 h
b.	Inexperienced draughtsman	4 h 40 min
	Editing of tracing process	1 h 30 min
	Average time per sq. in. of map sheet	a. 1 min
		b. 1.1 min

Scribing was carried out at the same time as tracing; the scribing time for neat line and present land use boundaries was as follows:

a.	Experienced draughtsman	3 h 31 min
	Editing	35 min
	Average time per sq. in. of map sheet	0.7 min

The same scribing procedure was also used to produce a copy of the present land use map of the Riding Mountain area, with its low-density data. The sheet measured 458 sq. in. and required 5 hours 12 min for scribing and 47 min for editing. Average time per sq. in. of the map sheet scribed by an experienced draughtsman was 0.8 min.

It was concluded that to transfer data from the storage file by manual tracing processes, scribing was faster and more accurate than the traditional

method of ink or pencil tracing. This quicker method of manual data transfer was taken as a standard for estimation of the time required for the tracing effort contemplated in the Canada Land Inventory. In the experiments, scribing resulted in times of between 0.7 and 1.6 min per sq. in. for the type and density of data contained in the sample file of graphic storage, or an estimated labour time of between 7 and 16 hours to complete a sheet of 600 sq. in. A set of 500 maps would involve between 3,500 and 8,000 man-hours (19.9 and 45.5 man-months). More than 14,000 manuscript map sheets have been generated by the Canada Land Inventory. Although the instruments and equipment necessary for this type of work are limited to normal draughting facilities (quiet room, tracing tables, scribe points, and scribecoat materials at a cost of approximately \$5.00 per sheet), it can be seen even from this inadequate sample that the requirements of labour and, more importantly, of the time to transfer the data, are large.

A preliminary investigation of human tracing error was conducted jointly with the timing studies. The sample was a land resources map of average data density; though subjectively chosen, it was representative of many in the Canada Land Inventory. A master copy of the map was reproduced photographically on stable material. Draughtsmen and operators with various degrees of experience were selected. Each traced the master copy by pen on a translucent stable material, first at normal speed, next at high speed, and finally at slow speed with maximum accuracy. After all tracing had been completed, a small area of the master copy was selected and enlarged on a high-precision enlarger. The same area on each of the traced samples was similarly enlarged, and deviations from the original more than one line wide were measured. As the data extracted were the boundaries of present land use areas, and as the factor critical to subsequent decision making was the measurement of size of the area traced, the tracing error was expressed in terms of its effect on the size of the areas subsequently measured. Based on the data recorded, it was found that the error was closely related to boundary length of the specific area and could be expressed as  $0.6 \times 10^{-3}$  sq. in. of error per linear inch of boundary measured. Thus, on a map at 1:50,000 scale, the average error incurred would be  $\pm 3$  acres in 1,300 acres. Although the time required is substantial, the tracing

method of graphic data transfer is therefore sufficiently precise to be adequate for a wide variety of decision-making purposes.

Measurement and comparison. Typical operations involved in extracting data inherent in a graphic file structure are counting of data elements, determination of distances and directions, and calculation of the size of areas. Comparisons of data elements, or of the results of their measurement, allow identification and extraction of such relationships as nearest neighbours among data elements, shortest routes among alternatives, areas with specific characteristics (such as size), and lines of sight. Measurement and comparison according to more complex sets of rules permit contextual search (identification of the occurrence of specific data elements in specific juxtaposition), and its sophisticated case, pattern recognition. Further extension of this operation allows the identification and extraction of groups of patterns occurring as spatially discrete sets, and the measurement of spatial correlations between elements or sets of elements that interact in the same space. All manual data extraction processes are undertaken under the mental control (monitor) of logical requirements prescribed by the decision-making process. This function is taken for granted when human skills are used but represents an extremely high degree of sophistication in computer-aided techniques.

The differences between the results of estimation and physical measurement in carrying out similar functions lie in their precision and the effort required to achieve them. The need for precision should, of course, be determined by the intended use of data resulting from the operation. Precision can usefully be thought of as "repeatability". It is an objectively determined level of activity at which it is decided to work, appropriate to the aims of the exercise being carried out. A related but different property, that of "resolution", is dependent on the size of the units utilized for measurement. The precision and resolution of data extraction clearly affect the subsequent use of the data and similarly influence the speed and level of effort needed to obtain them. In trivial operations, such as the counting of a small number of points, the differences are small whether such an operation is carried out mentally or with the aid of a simple mechanical counter. As the complexity of measurement and comparison increases, however, the

difference between estimation and physical measurement becomes significant and influences the uses to which the results can be put.

The determination of distance along a line may be used as an example. For simple shapes and lengths between 0.5 in. and approximately 30 in. an estimate of length usually results in errors of  $\pm 15\%$ . This level of accuracy is adequate for certain types of decision making. For intricate line shapes, particularly lines contained within a network, the error level increases sharply and estimation becomes less useful for the same decision-making process. Physical line measurement by various methods, such as comparison of line length with a scale of known length (either rigid or flexible), map measurement by moving wheel (curvimeter), or cursor travel measurement, can reduce the error to within  $\pm 4\%$  of the true measurement. Duplication of measurements can reduce the error further but the time required is increased substantially, frequently by factors of 10 or more.

An experiment was conducted in which 15 points were selected at random on a section of a map sheet that included a road network pattern. The shortest road distances between each point and every other point were first estimated and subsequently physically measured with a mechanical line measuring device. Findings are given in Table 3.1.

	<u>Time</u>	<u>Avg. % error</u>	<u>Standard deviation of % error</u>	<u>Range of % error</u>
Single measurement	220 min	Assumed to be zero in this case		
Careful estimate	50 min	9.3%	10.3	0 - 68.2%
Quick estimate	22 min	14.8%	12.0	0 - 64.8%

Table 3.1. Summary of results of distance estimation and measurement.

The average difference in error between the quick estimate and single measurement was approximately 15%, and the difference in time was in the order of  $\times 10$ . More significantly, the individual distances estimated showed errors of up to approximately 65%. Results obtained by measurement and estimation thus differ widely in their type of application to decision making, and require significantly different amounts of effort to acquire. It must also be remembered that this is a simple and straightforward example of data extraction from a graphic storage file.



A further example of data extraction is the overlay process, or measurement of the intersection of sets of homogeneous patches of areal data that occur in the same section of the spatial file.

Two types of physical measurement are in common use for manually measuring areas on maps, one using a polar planimeter, the other a dot area grid. Tests were done with both types of measurement. Part of the previously mentioned Greenville County soil map (1:50,000 scale) was overlaid with a transparent map at the same scale, and an area of 16 sq. in. was selected. The area contained 386 combinations of soil and present land use information, the size of the individual patches varied from 1 to 2,000 acres, and the total size of the area represented 10,240 acres on the ground. A standard equal-area dot grid was used because the inherent error in treating minimum-sized areas is smaller than that of a random dot pattern grid, which may miss many such areas and exhibit clustered point characteristics. The procedure illustrates the actions and the use of time involved in this type of measurement:

- a) A grid was selected; 1 dot = 1.6 acres.
- b) The grid was overlaid and fixed in position.
- c) The punch counter was used to record dots, a pencil was used to mark off measured areas. Where a dot was bisected by a line, a value of one-half dot was assigned to the area on each side of the line.
- d) The area of each combination of present land use and soils was measured and recorded separately in tabular form, as was the required time to complete the work on each area.
- e) Tabulation of the data was made on the basis of soil types and present land use. Summation of all combinations of area was made and the total compared with the known area.

The test was repeated using a standard polar planimeter with a tracing point, instead of the dot grid. The arm was adjusted to scale and readings were taken to three decimal places. Procedure:

- a) The operator traced around the area to be measured and recorded the reading in square inches from the Vernier scale.
- b) The operation was repeated three times, or until three readings differed by less than 0.003 sq. in.

- c) The average reading was recorded.
- d) Operation times were recorded as for the dot grid method.
- e) Area tabulations were made as for the dot grid method.

Comparison of results of tests on 16 sq. in. of map sheet are given in Table 3.2.

<u>Operation</u>	<u>Dot area grid</u>	<u>Polar planimeter</u>
Time to measure areas	9 h 6 min	14 h 36 min
Time for editing and tabulation	<u>3 h</u>	<u>3 h</u>
Total time	12 h 6 min	17 h 36 min
Average time per sq. in. of map sheet	45 min	65 min
Difference between sum of patch sizes and total size of area being measured	$\pm 2\%$	$\pm 3.7\%$

Table 3.2. Summary of results of area measurement test.

The results show that the dot grid provides a considerably faster and somewhat more accurate method of measuring map areas.

Similar area measurements were obtained from the Riding Mountain map sheet. The scale of map and overlay was 1:126,720, and that of the dot area grid was 1 dot = 13 acres. Two hundred square inches of map sheet were measured, edited, and tabulated (447,000 acres) at the rate of about 5 min per sq. in. This rate reflects two things: the relative coarseness of the grid and the low information density of the data in the graphic storage file. Percentage error was slightly more than 3% for the total area.

Measurement of low-density map data using a coarse dot grid thus required approximately 5 min per sq. in. and measurement of normal-density map data using a fine dot grid utilized 45 min per sq. in. Although the dot area grid may provide the faster and more accurate method of measuring map areas, the best time for a square inch of manual measurement apparently falls between 5 and 45 min for the comparison of two types of data. This does not include the preparation of the maps, but does include the measuring, editing, and written tabulation of the results.

The above tests were carried out because measurement and comparisons between maps containing different types of data were desired as an aid to understanding the quality and various characteristics of the agricultural

land in Canada. Given an average map of 600 sq. in. (20 in. x 30 in.) and assuming that an average of 400 sq. in. would be of interest for a particular examination, the time involved for the comparison of one set of 500 maps with another 500 would be between  $\frac{400 \times 500 \times 5}{176 \times 60} = 94.7$  man-months and  $\frac{400 \times 500 \times 45}{176 \times 60} = 852$  man-months. These times must be multiplied by the number of comparisons that need to be made, that is, the number of times it is wished to compare any two types of data throughout the inventory area. The small-scale, higher-speed approach results in the broadest spatial comparison of the data involved. If finer grids are used for measurement and comparison of more detailed data, the minimum times involved fall closer to the middle of the given range.

These figures cannot be extended into estimates of the total time and effort needed manually to extract data from a given set of maps unless one can decide, in advance, how many maps have to be examined and how many comparisons have to be made. Some maps might be expected to be the subject of numerous comparisons, but relatively few comparisons would need to be made on a countrywide basis. Nevertheless, the amount of time and effort needed to carry out manual measurement and comparison is evident and any large volume of maps would necessarily mean a lot of work. Equipment and instrumentation necessary for measuring and editing, however, are limited to normal draughting facilities, the provision of precision dot grids at approximately \$15 per set, or planimeters at a cost of \$65 each.

If measurement and comparison are replaced by estimation, less effort would be needed and more error could be expected. If other estimates of correlation between the data sets were applicable to the task facing the user, then sampling procedures based on fewer manual measurements and comparisons could yield useful results. In each case, a general level of precision is related to an approximate level of effort. In the example of measurement and comparison provided, the resulting data were needed to examine the actual relationship between present land use and soil characteristics in counties with varying socio-economic characteristics, as a basis for identifying marginal farming tracts in Canada.

## SUMMARY

This chapter has focussed on the utility of the map as a source of data from which information has to be extracted. General categories of manual techniques for handling mapped data have been examined and their efficiency has been illustrated with selected examples. Several conclusions can be drawn from the evidence presented.

From a simplistic, empirical point of view, it can be noted that for a single arbitrary area of about 10 sq. in. of map sheet, the various simple techniques of image manipulation, grouping, reading, search, selection, transfer, and measurement by estimation each take approximately 1 to 30 min to accomplish. There is wide variation within that range depending on the approach used, the density of data in the file, and so forth. For the operations of measurement and comparison, however, the time and effort required rise sharply to between approximately 80 and 650 min for the single 10 sq. in. This is illustrated on the following graph (Fig. 3.1).

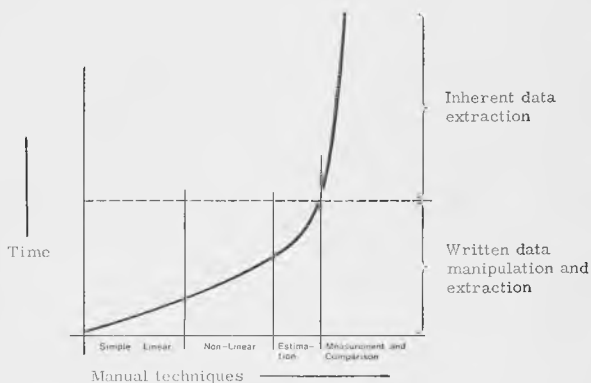


Fig. 3.1. Relationship between type of manual technique and time used.

Two comments may be made about the relationship shown. First, most techniques that fall below the break in slope (approximately 3 min per sq. in.) seem to be those that yield a reasonable return for effort. They probably

are currently considered to be economically practical as manual working tools. Secondly, these techniques operate on the data written on the map. Operations to extract data inherent in the file structure of the map fall above the break in slope and rapidly become uneconomical. It can reasonably be concluded that if manual techniques are used, graphic source data have low utility when data inherent in the file structure have to be extracted.

The above observations were derived from the time and effort involved in handling theoretically a single small piece of graphic source data. From the evidence in the text it is reasonable to assume that time and effort increase in proportion to the volume of graphic data to be handled. This may not be a linear function, as increasing data volume brings physical handling problems in addition to the time required for data manipulation and extraction. However, for the purposes of this argument the relationships can be considered to be linear ( $y = x$ ), as shown in the diagram (Fig. 3.2).

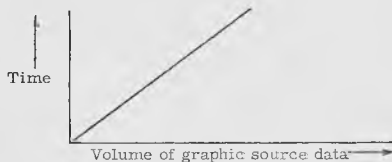


Fig. 3.2. Relationship between data volume and time required for data manipulation and extraction.

When the relationship between manual techniques and time is considered, the curve rapidly steepens ( $y = ax$ ).

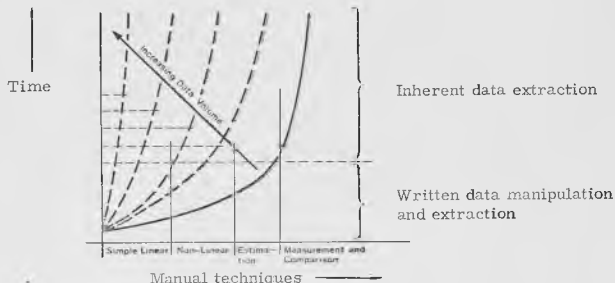


Fig. 3.3. Relationships between manual techniques, volume of data, and time used.

It can easily be seen (Fig. 3.3) that increasing data volume rapidly decreases the type of manual technique that can be considered economical. It also illustrates the point that increasing data volume rapidly makes uneconomical the manipulation and extraction of data actually written in the file. The effect of increasing data volume is particularly significant, as increases in file flexibility in a graphic medium usually imply creation of multiple graphic records and hence increase the bits of data to be handled. Graphic source material thus imposes an inherent inflexibility on file structures, and has low utility as a source of large volumes of spatial data.

From a theoretical point of view, it could be said that manual techniques for handling mapped data are reliant partly on manual motor processes, partly on mental data processing, and partly on human channel capacity. From the examples discussed in this chapter, it might be reasoned that manual motor processes are relatively slow, that mental graphic data processing is relatively fast, and that human channel capacity is apparently limited. In general, then, techniques that rely mainly on manual motor processes, such as data transfer by tracing or manual measurement, are slow. Techniques that rely mainly on mental data processing are fast; for example, estimation of line length is approximately 10 times faster than manual measurement of the same lines. Techniques that rely on combinations of manual motor processes and mental data processing increase in efficiency with the proportion of mental data processing involved, until human channel capacity is reached.

Two examples can be used to illustrate this last point. The first is centroid allocation, an operation combining mental data processing and manual motor processes. This technique is economically acceptable until very complex areas are encountered (such as a matrix in which many islands are studded), at which point mental retrieval of the image is so complex that channel capacity is exceeded and the error involved in centroid placement becomes very high. The second example is estimation of line length, which operates at rates of approximately 15% error when lines are short, but which rapidly increases in error rate for long complex lines. These relationships can be expressed diagrammatically as follows (Fig. 3.4).

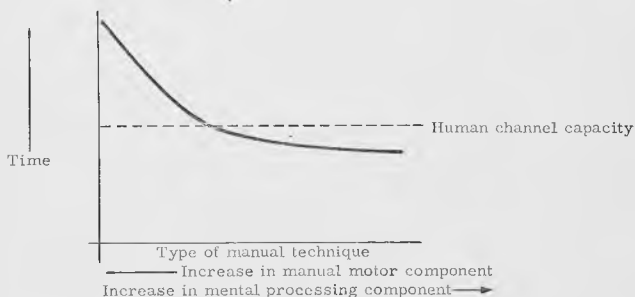


Fig. 3.4. Relationship between components of manual techniques and time used.

It must be emphasized that no measures of human channel capacity were undertaken during the work on which this chapter is based, and therefore these remarks are essentially speculative. From the empirical evidence that has been presented, however, it may be reasonable to conclude that if computer-aided techniques are to improve on manual techniques for handling mapped data, such improvement might profitably be sought in operations that reduce manual motor procedures and that allow the presentation of data to a user in a form that does not overload his channel capacity. An examination of current computer-aided techniques for handling mapped data is presented in the next chapter.

#### CHAPTER 4 - CURRENT COMPUTER-AIDED SYSTEMS FOR HANDLING MAPPED AND OTHER LOCATION-SPECIFIC DATA

Computer-aided techniques for handling mapped and other location-specific data may be described as current if they are beyond the development stage and have been implemented in government or academic systems for handling spatial data, for purposes other than the development of the techniques themselves. A spatial data handling "system" in this sense is taken to be a set of one or more computer-aided techniques for reading, storing, manipulating, and displaying spatial data. Such systems have been devised mainly in the past decade, and the field continues to develop rapidly. To understand the contribution of the current systems, it is necessary to examine them in the context of their development to date.

Before 1960, computers were essentially high-speed calculating machines. Large amounts of data could be manipulated in the computers, but their limited internal storage capacity was a significant constraint on their use. The early to mid-1960s saw the development of computers with very large magnetic storage capacities, in excess of one million characters. This development allowed rapid growth in the use of computers for information handling and, in particular, opened the door to the non-graphic storage of spatial data.

The computer storage of spatial data has two advantages. The first is the relatively compact nature of the storage compared with conventional graphic storage. The second advantage is more crucial, since it adds a new dimension to the first; this is the ability to use established computing machines to read, store, and manipulate non-graphic stores of spatial data. In addition, such stores of data are relatively easy to augment, edit, and correct without renewing the entire storage.

An obvious use of non-graphic storage is to collect together the numerical results of the manual processes of topographic survey, in preparation for various forms of graphic output. Advances in this area stemmed from the work of photogrammetrists with computers in the 1950s, and systems of instruments now exist that allow measurements from photographs to be transferred digitally to magnetic tape (automatic photogrammetry). The



use of non-graphic storage by traditional map-making institutions (automatic cartography), however, has grown only more recently. Existing systems can now take numerical survey data as their direct input and process it digitally to produce a non-graphic store of spatial data; this can easily be displayed, updated, and edited. Also, plotting machines can now accurately convert a non-graphic store into a graphic display. These systems of automatic cartography, however, are still in their infancy or, at best, at an experimental stage. At present (1974), no extensive non-graphic store of topographic information has been generated directly from numerical survey data.

The existence of large and growing stores of spatial data in graphic form, and the ability of present-day computers to manipulate non-graphic storage, have led to the development of techniques to convert graphic to non-graphic formats and to manipulate the resulting stores of non-graphic spatial data. Several methods now exist for such transformation. Once a non-graphic file of spatial data has been created, data may be added to it without passing through a graphic stage. Various non-graphic data formats and related file structures have been developed, with differing degrees of resolution and ability to handle the original data. Early efforts in such transformation employed simple data formats and file structures, several of which are still in widespread use (for example, grid systems). The general acceptance of these approaches should perhaps be questioned, bearing in mind the relationship between data availability and the decision-making process. More sophisticated non-graphic data formats and file structures have been developed and are becoming more widely used. The various manipulation facilities that have been developed rely to a considerable extent upon the file structures to which they are applied.

Although the processes of handling spatial data have undergone considerable development in the past decade, current techniques are in a transition phase. At present their data arrangements are based, with few exceptions, on local applications. The international adoption of common data formats and file structures is a trend which must be expected, however, and this standardization may be hastened by the need to monitor environmental

values on an international scale. It may be prescribed by traditional data gathering institutions. We shall also see the development of systems that can accept numerical data from survey instruments to augment established non-graphic data bases. Within the next decade also, computers that use optical, holographic, or advanced forms of electromagnetic storage will probably become available, allowing noncoded (for example, graphic) data to be stored and manipulated. This will substantially reduce the need to convert graphic source data to coded non-graphic formats and, with the advent of economical picture processing, will measurably increase the range of graphic data types that can be processed. The trend of these developments is shown schematically (Fig. 4.1).

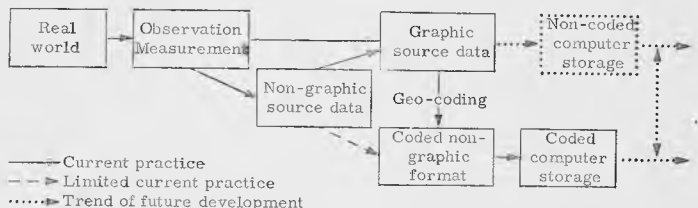


Fig. 4.1. Developments in data flow.

Before future trends are discussed, however, it is the purpose of this chapter to examine currently developed techniques for handling mapped data and to consider the utility<sup>1</sup> of computer storage as a source of data from which information has to be extracted.

#### BASIC CONSIDERATIONS

For all computers in general use in 1974, mapped and related location-specific data are retained in a coded, non-graphic form of storage. The

1. The "utility" of a store of information for decision-making purposes is dependent upon 1) whether the types of data in the store are relevant to the decisions being made, 2) whether the resolution of the data elements (and by implication their format) is relevant to the decisions being made, and 3) how easily the data in the store can be manipulated (combined, correlated, compared, etc.) to produce the information on which the decisions can be made.

The relationship between stores of information and decision-making processes will be examined later in the thesis. In this chapter attention will be focussed on data handling techniques from a functional standpoint. The "utility" of computer storage in this sense refers to the limitations and capabilities inherent in non-graphic storage for recording and manipulating graphic source data.

character and the utility of that storage are functions of the type of non-graphic data format or formats used and the choice of file structure employed (Fig. 4.2).

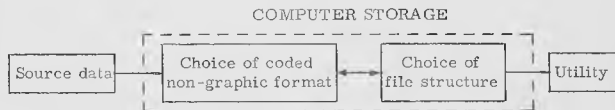


Fig. 4.2. Components of computer storage.

In very general terms, the choice of coded data format determines what subsequently can be done with the data, and the choice of file structure determines the relative ease of the various manipulative possibilities. These two aspects of computer storage are obviously closely interrelated. Of the two, the choice of coded format is perhaps the most critical as it can have impact on the content of the data. Transformations of data to various coded formats, or from one coded format to another, in certain circumstances can result in information loss. This is examined more fully below. Transformations between file structures, on the other hand, are functions of time and cost, these in turn being related to the capabilities of the computer being used. The type of file structure used will be (or should be) guided by the intended use of the data, whereas its complexity is directly dependent on the coded non-graphic format and the volume of data to be handled.

#### CODES FOR NON-GRAPHIC SPATIAL DATA

Two broad categories of codes for non-graphic storage of spatial data can be recognized. The first of these is metric coding, based on a geometrical coordinate system; if a high degree of accuracy in geographical locations is required, then some form of metric coding must be employed. The second is topological coding, which defines the locations of data elements in space with respect to one another, but without reference to actual distances. Quite rigorous topological grammars can be developed to define spatial relationships in ways that greatly facilitate data editing, data manipulation, and data retrieval for specific purposes. Either category of

coding can be applied to a given set of spatial data, but powerful capabilities for data manipulation can be developed when both approaches are used in combination on the same data set.

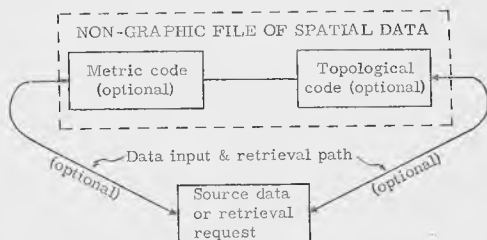


Fig. 4.3. Data input and retrieval paths.

The diagram of data input and retrieval paths (Fig. 4.3) illustrates the last point. A given set of source data may be coded metrically, topologically, or both. If both methods are employed, a request for retrieval of source data may enter the file via the topological codes, which can be used as an index to the metric data, or via the metric codes, which can be used as an index to the topologically coded data; the choice depends on the nature of the enquiry. In addition, topological codes can be used in conjunction with metric codes to reduce the amount of overall storage required.

All metric codes rely either implicitly or explicitly on a coordinate system. As mentioned in Chapter 3, at present the only widely accepted model of the space occupied by the earth is Euclidean. The primary category of non-graphic storage would seem to be the numerical record of locations in terms of their three-dimensional spatial coordinates. However, relatively complex calculations are required to manipulate this type of coordinate system, the interpolation of data elements is also complex, and there are currently few government decision-making processes that require data to be manipulated on a three-dimensional global basis; this is therefore an uncommon form of non-graphic storage at the moment. Future development of global environmental monitoring systems may require transformation of some types of data to such forms of non-graphic storage.

Plane graphic representations (maps) of the curved surface of the earth are devised by mathematical transformations. The location of data elements mapped on the plane can thus be numerically coded by map coordinates. Similarly, data can be coded in terms of an arbitrary system of plane coordinates superimposed on the map. Such arbitrary coding is used, for example, in conjunction with an automatic tracing input device; or to aid subsequent computation within the computer; or to help transform arbitrary coordinate values to the original plane coordinate system or to another system.

In a plane coordinate system, the coordinates can be used alone, to define the location on the plane of points (of places or events), arbitrarily shaped areas (such as grids), lines of any configuration and, by extension, area boundaries of any configuration. The spatial juxtaposition of such data elements is made explicit in their coordinate definitions.

At the other end of the range of non-graphic data formats is the exclusive use of topological codes to record spatial data elements. The simplest form is a nominal code which identifies data as belonging to a particular geographic area or location. This value does not convey relative location directly, but must be used in conjunction with an external metric index, typically a map. Common among such nominal codes are numbers or names identifying administrative areas, street addresses, census tracts, traffic zones, and so forth. Most existing geographically referenced data that are not mapped employ this descriptive coding method.

Still at the exclusively topological end of the range is the use of codes that specify the relative location of data elements (for example, point connectivities, grid cell contiguities, networks, and area adjacencies) but where the actual location of the data elements in space must still be determined from an external index. A useful example of this approach is a basic DIME<sup>1</sup> file which provides a topologically rigorous connectivity code for city streets without necessarily mapping the resulting network in space.

Between the extremes of exclusively metric or exclusively topological

1. DIME is the Dual Independent Map Encoding system (U. S. Bureau of Census 1972).

coding is the combined use of the two types of code (Fig. 4.4).

Only metric codes used*	Metric and topological codes used*		Only topological codes used*
Pure coordinate definition	Data compaction functions	Data connectivity functions	Names referring to external metric index
	Decrease in use of metric codes		
	Decrease in use of topological codes		

\*"Used" in this sense means that the code or codes concerned are part of the storage of the data handling system and the properties they describe are thus known to the system.

Fig. 4.4. Relationship between metric and topological codes.

When the metric and topological properties of the data set are known to the system to varying degrees, various types of compaction of data storage or indications of data connectivity, or both, are possible, though at the expense of complexity of the file structure (see discussion of file structures below). This can be illustrated with examples for each of the four categories of metric code (points, arbitrary regular areas, line segments, and irregular polygons).

Points defined by coordinates can be given names (typically, numbers) which can be used in subsequent grouping processes, thus avoiding the repetition of their lengthy coordinates; this is a simple form of data compaction. Point names can be related to one another by a point connectivity code or a point connectivity matrix. Points can be related to other data elements, whose locations can be either defined or undefined, by either connectivity codes or a connectivity matrix. A case in point is the use of area centroids as an index to areas. If the areas are defined by coordinates, the coordinates of the centroids can relate them to the areas. Coordinate points can also be related to areas by implicit or explicit connectivity codes. In either case, the metric relationship of the points can be used (and frequently is used) as a surrogate for the metric relationship of the areas, with some gain in data compaction but with the risk of loss of accuracy. In addition, centroids with metric or other codes can be

topologically interrelated, for example by a point connectivity code or matrix, which may function as a surrogate for an area connectivity indicator, but again with the risk of loss of accuracy.

Arbitrary regular areas can be defined by coordinates fixing two or more points of the array and by topologically coding the regular cells contained in the array (a form of data compaction). Both sequential data grouping and inverted data grouping of areal units similar in size and shape are common (see discussion on file structures below). Explicit data element connectivity is inherent in both forms of topological coding.

Line segments can be recorded in a compact form of storage that utilizes a point-of-origin coordinate followed by their length and direction. With a polar coordinate system, an approach using variable-length vectors can be employed. With an orthogonal coordinate system, a chain-encoding approach can be used to record changes of direction from a starting point; it is not necessary to specify length in this case, as it is implicit in the direction codes because of their hierarchy of movement. This approach to line segment compaction coding was first employed in CGIS.<sup>1</sup> Line segments can be named, and connectivity between line segments and other data elements can be separately indicated, for example, by coding to end points or to contiguous areas.

Irregular polygons (areas) can be named and the names used as surrogates in subsequent grouping processes (a simple form of data compaction). Areas can be encoded as polygons made up of a series of points, or as a set of line segments; this is the redundant converse of line segments coded contiguously to areas. Areas can be related to other areas by being coded as contiguous or by a connectivity matrix.

These examples do not form a complete list of the specific non-graphic data formats that can be devised. They are given to illustrate the range of possible types of code and the range of types of spatial relationship within a data set that can be expressed by various coding forms. Examples of

1. The Canada Geographic Information System (Tomlinson, R. F. 1967).

the possibilities are given below (Fig. 4.5).

Pure coordinate definition	Examples of data compaction functions (Combined use of metric & topological codes)	Examples of data connectivity functions	Names referring to external metric index
Points	Point name (typically a number)	Point to point connectivity Points related to areas	<div> <div>code or matrix</div> <div>code or matrix</div> </div> <div> <div>Topologically unrelated names</div> <div>Topologically related names</div> </div>
Arbitrary regular areas	Grid cell name Sequential coding Inverted coding, etc.		
Line segments	Segment name Variable vector coding Unit vector coding (chain coding)	Line segments to end points Line segments to areas	
Irregular polygons (areas)	Area name Line segment boundary chains	Areas made up of points Areas made up of line segments Areas to areas	

Fig. 4.5. Range of coding related to expression of spatial relationships within a data set.

It can be seen from the diagram that the basic format of non-graphic data is decided by the type of real-world data element being encoded (point, line, area, or arbitrary area). The simplest form of non-graphic storage uses pure coordinate definition alone, or topological reference to an external index. The central part of the diagram, however, indicates the area where careful choice of file structure can benefit the subsequent ease of retrieval of the data, can allow modal change of data to be attempted, and, in some instances, can reduce storage requirements.

#### FILE STRUCTURES FOR CODED NON-GRAPHIC SPATIAL DATA

The physical arrangement of any substantial collection of data in any format, non-graphic or graphic, affects the ease with which it may subsequently be examined, correlated, and understood. The range of data arrangement types that can be employed is directly related to the physical and economic constraints of handling data in the chosen storage medium. The choice of data arrangement within that range is (or should be) based on the manner in



which the data are subsequently to be manipulated.

The physical characteristics of storage media readable by a computer (such as cards, tapes, disks, core, and cells) and the electronic capability of existing computers to read and handle data so stored are well-documented aspects of computer design. The resulting range of data arrangement types is discussed briefly below so that the file structures<sup>1</sup> currently used to handle mapped and related location-specific data can be viewed in context.

The time and cost involved in reading and correlating a collection of computer-stored data are affected by: the sequence in which the data are stored (file sequencing); the manner of linking and the degree to which it is possible to establish links between data or sets of data (file linkage); the manner in which each data element is described in the file (file record layout); and the information content of the file (file information content). Only a limited series of approaches can be used within each of these four aspects of file structure. Although there is no accepted ordinal scale of complexity of overall file structure, the simplest file structures tend to use elementary approaches in these four categories and the most complex file structures employ one or more of the advanced approaches. A brief description of the various approaches follows.

#### File Sequencing

- 1) Sequential
- 2) Indexed Sequential
- 3) Direct Access

A sequential arrangement is the simplest of all. Each data element has a direct and specific relationship to the ones stored immediately before and after it, and this relationship is specified by the position of the data element in the sequence. It may be argued that the lowest form of sequential arrangement is one where the data elements deliberately have no mathematical relationship to each other (that is, nominal scaling). The various forms of ordinal scaling (partial ordering, weak ordering, and complete ordering), interval scaling, and ratio scaling form increasingly sophisticated levels of sequential relationships. Multidimensional scaling cannot be conveniently accommodated in a simple linear arrangement without transformation. If an array of grid squares were sequentially filed, they would be stored strip by strip.

1. For the purposes of this discussion, a "file" is taken to be an arranged set of data, and file "structure" refers to the manner in which the data are arranged.

An indexed sequential arrangement implies a sequential storage relationship between data elements, but adds the facility of a related index. If, for example, some data elements in the sequence were missing, the relationship between adjacent elements in the file could be determined from the index. For example, if an array of grid squares were indexed by grid centroid coordinates, the result would be an indexed sequential file.

A direct access arrangement does not require any relationship to exist between adjacent data elements in storage (an argument against nominal scaling as the lowest form of sequential arrangement). In this case, however, the data elements are placed in a logical sequence according to a randomizing formula and the relationship between data elements is determined entirely from the randomizing formula and not from their position in the sequence of data elements. A common use of a randomizing formula is known as "key transformation" or "hash coding". The name of each data element to be stored, for example, the value of variable A at grid location 23, 64 (i.e., A, 23, 64) is transformed into a number calculated from the binary form of the expression A, 23, 64, using a given formula. The resulting number is used as an "address" in computer storage at which are stored the data describing the actual value of A at grid location 23, 64. Even with some additional codes to allow for reassignment of duplicated addresses, sparse data contents of an array can be stored in much less space than would be taken by a complete array, and the addresses of specific data items can be determined through calculation by the randomizing formula.

#### File Linkage

- |                                                                |                                |
|----------------------------------------------------------------|--------------------------------|
| A) <u>File data organization</u>                               | B) <u>File pointers</u>        |
| 1) <u>Sequential data grouping</u>                             | 1) None                        |
| 2) Inverted data grouping                                      | 2) Intra-data-element pointers |
|                                                                | a) External directory          |
|                                                                | b) Internal directory          |
|                                                                | c) Integrated pointers         |
| 3) List-structured data grouping                               | 3) Intra-file pointers         |
|                                                                | a) External directory          |
|                                                                | b) Internal directory          |
|                                                                | c) Integrated pointers         |
| 4) Value-list-structured data or<br>Cross-list-structured data | 4) Inter-file pointers         |
|                                                                | a) External directory          |
|                                                                | b) Internal directory          |
|                                                                | c) Integrated pointers         |

The purpose of file linkage techniques is to allow or aid the grouping and retrieval of subsets of data contained in a file, or in multi-file collections of data, without reading and processing the entire data store. One obvious approach to this task is to determine in advance all the needed subsets of data and to create separate small files for them. This is the basic concept behind file data organization.

Another approach is to include in the original file or files a set of "pointers"<sup>1</sup> leading from appropriate data elements to other data elements with which they may eventually be grouped. This is the basic rationale behind file pointers. Both approaches can be used within the same data set.

The simplest form of file data organization is sequential data grouping. No subset grouping is attempted, and the only subsets that can be conveniently extracted are parts of the original file in their original sequence, such as the first half, the middle third, or the last quarter. Inverted data grouping is the first attempt at reorganizing the original file. It is the elementary concept of matrix inversion. Data in an array of grid squares would be stored column by column rather than row by row, to avoid processing the whole file in order to examine the data from one column only. This straightforward approach becomes expensive if the data contained in the array are sparse. To overcome this problem, list structures are used which simply list the data in the subsets required and store them in that manner. A value-list structure is employed when each data record contains a number of values, and subsets of selected values are required as well as subsets of the data records that contain those values. Similarly, lists can be created that group specific data elements from a variety of other lists; this provides a cross-list structure.

All of these forms of file data organization can be achieved by using pointers; in fact, list structures (A3 and A4 in the subsection heading above) are not often found without them. Pointers can be contained in a list that is external to the system (external directory), or a list that is internal to the system and hence part of the file (internal directory), or they can be actually attached to the data elements themselves (integrated pointers). Intra-data-element pointers are those employed to link a series of items

1. Pointers can take several forms, the most common of which is a code followed by the storage "address" of the next desired data element.

of data that together make up one data element (for example, the linkages between line segments that together make up a specific polygon). Intra-file pointers are those employed to connect subsets of data within one file (such as the linkages that establish adjacency between polygons on a map). Inter-file pointers are those used for multi-file indexing (for example, to connect different data types). The last category of pointer permits more efficient use of subsections of data for specific purposes and more efficient use of some types of computer hardware that handle small bundles of data as cheaply as individual data elements.

The sensible use of various types of file sequencing and file linkages increases the flexibility of a file of data and decreases the time for data retrieval in response to an enquiry. These benefits are gained at the expense of using additional storage space and increasing the time needed initially to establish and subsequently to amend the file.

#### File Record Layout

- 1) Fixed Format
- 2) Variable Format
- 3) Mixed Format

Practically all computer storage systems are linear; one storage location comes after another and no new ones may be inserted between the original ones. The simplest type of record layout is thus one where the format of each data element can be determined in advance and where fixed blocks or standard-sized units of storage are allocated for each data element. Such records are said to be in a fixed format.

Apart from the obvious advantages of uniformity, the value of fixed-format data is apparent in all of the data structures described above, where the position of a datum in a file carries a great deal of meaning and can identify the datum. The difficulty arises when data in a fixed format structure are sparse or when, by the nature of the data, a fixed format is impracticable. Areal data with boundaries described by line segments is just one example. For such cases, variable-format data can be employed, but additional data (pointers) are needed to indicate the beginning, size, or end of a record, or to indicate the size of variable-sized data arrays. Where the data content of each record is variable (variable data contents), a directory of

the contents must be provided in a separate part of the computer or coded within each record itself, so that the various categories of data at different places in the variable records can be located, added to, or deleted. Mixed formats may be used in large files, to take advantage of fixed formats for some parts of grouping and retrieval while retaining variable formats to accommodate variable forms of data. Inevitably, the interaction of the two requires additional sets of pointers, storage space, and programming.

#### File Information Content

The structure of any file is obviously influenced by the format of the source data it has to handle (see coded non-graphic data formats, above). A second influence on the utility of the file structure is its capacity to accept related or summary data which make it unnecessary to repeat calculations on the original data set (for example, the area or the centroid of polygons).

The sensible use of file record layout and additional file information content can decrease the volume of storage required for sparse data, at a cost of increasing the complexity of the file and hence the processing costs. (Where data subsequently become dense, both storage and processing costs are increased.) The task that faces the file designer is to achieve a balance between storage volume and the amount of processing needed.

#### RELATIONSHIP BETWEEN TYPE OF STORAGE FORMAT FOR NON-GRAPHIC SPATIAL DATA AND CURRENT SYSTEMS FOR HANDLING SPATIAL DATA

The underlying theme of the above discussion is that the format, file structure, and utility of a store of non-graphic data depend to a great extent on the format of the real-world data elements to be represented in it. When various formats of real-world data must be included in one system, it is, of course, possible to code them and combine them in a variety of ways. When there is only one type of format of real-world or coded data element, there are inherent limitations on the types of operations that can be performed on the resulting store of data. This can be illustrated with point data.

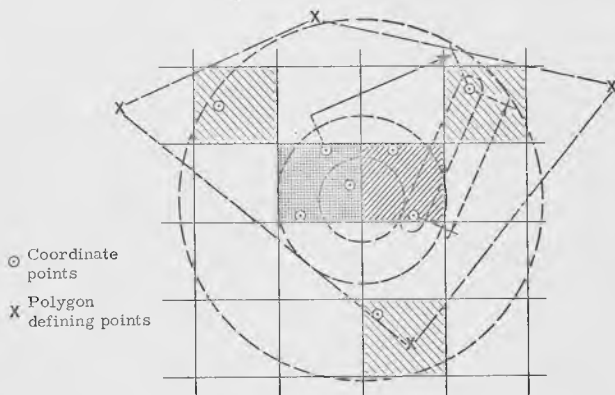


Fig. 4.6. Operations possible on spatial data coded as coordinate points.

The diagram (Fig. 4.6) shows the operations possible on a non-graphic store of spatial data coded as coordinate points. The points can be aggregated into coarser grids, calculations can be carried out on density (including contouring), distance, area, and direction, and boundary operations can be performed (for example, within circles, rectangles, or polygons). A basic limitation of the format of non-graphic storage using nothing but coordinate points is that only relationships derived from point-to-point distance calculations can be adequately rendered. Without additional data derived from other data formats, there is no way of inferring information about adjacency or accessibility such as could be determined from network, grid, or boundary (web) formats; these are not inherent in the point system.

The type of limitation applies to some degree to all individual data formats, and has particular impact on the ability to transform one type of data to another. For each type of real-world data element there is still an optimum coded non-graphic format (points as points, lines as lines, and so forth). Transformation to other formats is possible in some instances, but in others it requires data reduction. In this early stage of development of computer-

aided systems for handling spatial data, there has been a strong (and perhaps understandable) tendency to avoid the complications of cross-format linkages but rather to develop systems primarily designed to handle specific, individual formats of data. The relationship between real-world data elements and coded non-graphic storage formats with examples of current systems is shown in the diagram (Fig. 4. 7).

Real-world data element Coded storage format	Point	Line	Area	Typical current geographical information systems
Arbitrary regular area (e. g., grid)	REDUCTION OF INFORMATION			LUNR, OZARKS, etc. SYMAP, MIADS, etc.
Point				FRIS, GRDSR, etc.
Line segment	POSSIBLE INTEGRATION			DIME, SACS, etc.
Polygon				MAP/MODEL, PIOS CGIS

Fig. 4. 7. Relationship between real-world data elements and coded non-graphic storage formats.

Each of these categories of storage format, that is, arbitrary regular areas, points, line segments, and polygons, will be examined in turn below. Their characteristics will be illustrated by examining one or more specific systems within each category, so that the practical, as well as the theoretical, constraints and capabilities inherent in each storage format can be seen.

1. LUNR - New York State Land Use and Natural Resources system; MIADS - U. S. Forestry Service system; SYMAP - Synagraphic Computer Graphics, Harvard University; FRIS - Swedish Central Board for Real Estate Data system; GRDSR - Geographically Referenced Data Storage and Retrieval system, Statistics Canada; DIME - Dual Independent Map Encoding system, U. S. Bureau of Census; SACS - Street Address Conversion System, Urban Data Center, Seattle.

#### ARBITRARY REGULAR-AREA CATEGORY OF SYSTEMS

It can be seen from the diagram (Fig. 4.7) that the choice of a storage format using arbitrary regular areas (typically, a grid) for real-world data elements implies some degree of data reduction, unless the source data elements are homogeneously (evenly) distributed throughout each arbitrary area and unless changes in data values occur only at the boundaries of the arbitrary areas. The method has the advantage of being a simple approach to data storage, at the expense of some information reduction. Although considerable (usually manual) effort is involved in reworking existing data sets to conform to regular-shaped "pigeonholes", it is a simple computational task to store the results and subsequently retrieve or display them.

If the arbitrary areas are large compared with the size and spatial complexity of the data elements they represent, the coding frequently creates a spurious homogeneity, degrades locational precision, and imposes an unreal format on the source data. If the arbitrary areas are small compared with the data elements they represent, these effects are minimized. A wide variety of systems has been developed using the arbitrary regular-area storage format, and they range from those initially designed to accept comparatively large cells relative to the data being stored, which will be termed Type 1 systems, to systems whose manipulative functions assume a small cell size relative to the source data elements being stored, to be termed Type 2 systems (Fig. 4.8).

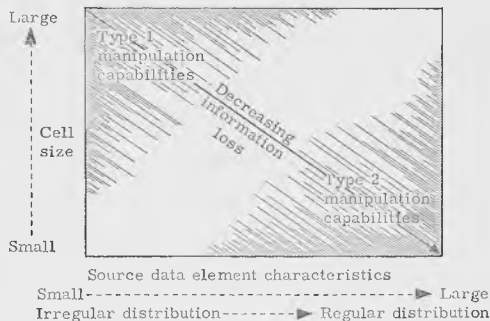


Fig. 4.8. Types of manipulation capability in systems based on storage formats using arbitrary regular areas as a function of cell size and characteristics of source data elements.



Inherent information loss generally decreases along the diagonal in the diagram (Fig. 4.8). Within specific systems, however, the degree of data reduction is a function of the mixture of characteristics of the data elements being stored. The relationship between manipulative functions and the ratio of cell size to data element size is simplified, but this allows a general grouping into two broad categories of systems for descriptive purposes.

#### Arbitrary Regular-Area Systems - Type 1

The first category (Type 1 in Fig. 4.8) is a type of data structure that has been adopted for a number of administrative units in various countries. In North America, for example, the type of storage format using arbitrary regular areas is usually a grid. The size of the grid square employed varies from 1/8 mile square in Illinois (NARIS system<sup>1</sup>) and Minnesota (Minnesota Lakeshore Development Study<sup>2</sup>) to 500 m square in parts of Ontario, Canada (Erie Project<sup>3</sup>), 1 km<sup>2</sup> in Massachusetts (Comprehensive Land Use Inventory Project<sup>4</sup>) and Delaware County, Pennsylvania (Delaware County Information System<sup>5</sup>), and 10 miles square in the Ozarks region (Composite Mapping System<sup>6</sup>). One typical example of the first category is LUNR,<sup>7</sup> the Land Use and Natural Resources inventory of New York State, which employs a 1-km<sup>2</sup> grid square. The LUNR system will be used to illustrate the problems and potentials of this category of storage format because it is a straightforward example of the genre which has been fully developed in terms of the original objectives of its designers. It is quite well documented and has been in existence for a sufficient time (since 1969) to allow a preliminary assessment of its capabilities.

#### LUNR System - Example of Type 1 of the Arbitrary Regular-Area Category

The aim of the LUNR<sup>8</sup> project was to produce an inventory of the present land uses and certain environmental line and point attributes in New York State. The data sources available to the project were aerial photographs (1:24,000) supplemented by existing maps, reports and directories, public agency records, and direct field observation. The data elements identified were plotted on 7-1/2 minute series of topographic base maps (1:24,000) of the United States Geological Survey (USGS). The first step

1. University of Illinois 1970. 2. Borchert 1969. 3. Bruer 1967, Philbrick 1967. 4. Vogt, Ivers and Associates 1966. 5. Delaware County Planning Commission 1970. 6. U.S. Dept. of Commerce 1967. 7. Hardy et al. 1971a. 8. Hardy et al. 1971b.

in the overall project was thus to establish a totally graphic storage file.

The graphic data are converted to a non-graphic format using very simple combined metric and topological codes for an array of grid squares<sup>1</sup> superimposed on the USGS base maps. Universal Transverse Mercator (UTM) coordinate pairs taken from the corners of each base map metrically define the array. The grid squares are topologically coded by sequential numbering (left to right, row by row) within each sheet. All data elements are then assigned to a unique grid cell number.

To assign data elements to grid cells, the total of areal units (indicating present land use) within each cell is manually measured by a type of dot grid accurate to 0.01 km<sup>2</sup> and expressed as the percentage of the cell covered by a particular type of areal data. The resulting measurements are coded and assigned to the grid square number. Linear measurements within each cell (mileages of shorelines, railways, and so forth) are performed to a resolution of 1/10 mile; these data, and counts of numbers of point and area data elements, are similarly assigned to grid cell numbers. Thus, a significant amount of data extraction (measurement) is carried out on the graphic storage file at this stage, and it is the results of these measurements that are stored in the regular matrix of the non-graphic storage file. No detailed cost breakdowns of manual effort are available, but it can be noted that 140,000 1-km<sup>2</sup> grid cells are needed to cover the State of New York, and the manual part of the project was large enough to demand the full-time efforts of up to 50 workers over a period of four years (1967-71).<sup>2</sup>

The coded non-graphic data are keypunched and stored on an IBM 2316 Disk Storage device.<sup>3</sup> Each individual record contains all the data for five grid cells, organized sequentially in 560 "words" (that is, 115 words per cell). The data are compacted on the basis of one word (4 bytes, 32 binary bits, or 24 floating point bits, plus 8 bits exponent) per set of 8 or 10 real numeric integers, and stored in fixed-format records; empty fields are filled with zeros. The records are filed for direct access, though the randomizing formula used creates 140 indexed sequential groups of records, each group containing the records from one base map.

1. It should be noted that the grid "squares" are only approximately square and their shape varies slightly over the extent of New York State. This inherent error was not considered to be significant in any calculation using the summarized grid cell contents. 2. Hardy and Shelton 1970. 3. Shelton et al. 1971.

The retrieval program (DATA<sup>1</sup>LIST<sup>1</sup>) can select any individual record or set of records. Any cell can be called<sup>2</sup> by using the coordinate point of its north-east corner; similarly, any rectangular block of cells can be called by the coordinate points of block corners. Cells can also be called by data items, which may include county and township names. On retrieval, each individual record is expanded into the full data set for the five separate grid cells. Unwanted cells or data values are simply discarded. This is the only form of retrieval possible, and it must be used to get even one value from one grid cell. The retrieval of the total data from one cell (or set of five cells) is thus relatively efficient, but the retrieval of one value from the entire file requires processing of the entire file. There are no explicit pointers in any form incorporated in the file. Users have created specific list structures for frequently required retrievals in selected areas, but these are not incorporated or indexed in the main file.

Any combination of data values from cells can be produced and any mathematical or logical analysis can subsequently be made on the data, although specific algorithms must be programmed for the data manipulations required in these tasks.

Using a second computer program (PLANMAP<sup>1</sup>), a pictorial representation of the grid squares can be produced by the computer line printer. In conjunction with DATA<sup>1</sup>LIST<sup>1</sup>, the mapping program PLANMAP can display any selection or mathematical combination of grid values. Symbols can be overprinted to give different grey tones (choice of five levels) in each grid. Numerical values (one digit) can be printed in the middle of each grid so displayed. A great many line-printer maps of the grid squares, showing different combinations of items, can be produced very quickly (within hours). While the content of the grid remains constant, it is possible to change the "scale" by increasing the area of the line-printer paper on which each grid is produced. No other image manipulations or types of data extraction are possible in this system.

To assess this type of system in the first category of storage formats, one could compare the ability of the LUNR system to manipulate its non-graphic storage with the ability to manipulate the original graphically stored data

1. Shelton and Dimock 1971a, 1971b. 2. "Called" in this context means selected for retrieval.

manually. The original graphically stored data could be assumed to comprise approximately 140 sheets of point data similarly recorded. The effort required to summarize and overlay the various types of data elements from the contrasting formats could be estimated and compared. Such a direct comparison would, however, be misleading, and therein lies the key to assessing the utility of LUNR-type systems. The non-graphic store does not contain the same data set as the graphic store. It contains a selected set of measurements manually derived from the graphic store at considerable expense and recorded in an arbitrary format. The LUNR system can only manipulate and display those measurements; it has no access to the source data.

As mentioned in the system description above, the manual process of creating non-graphic formats of arbitrary regular areas involves measuring and coding all present land use areas, measuring and coding line lengths, and counting and coding points. These are high-order manual techniques (see Chapter 3, p. 78) that extract data inherent in the file structure of the maps involved, and for reasons of cost and time they are not usually applied to more than a limited number of map sheets.

If these measurements are grouped according to their occurrence in  $1\text{-km}^2$  cells, the original spatial relationships of the data elements are lost and are replaced by the arbitrary relationship inherent in the regular area format. This information reduction includes the loss of original area boundaries (both their configuration and precision of location), loss of precision of point location, and loss of line segment configuration and precision of location. Underlying the entire manual transformation process is the assumption that the measurements being made will be required by the user and that the loss of information is not important to the user. In terms of the use of the subsequent data, the loss of information is, of course, greatest when small numbers of cells are considered, and is minimized when large numbers of the cells are considered.

The tasks eventually assigned to the computer to manipulate this non-graphic storage are relatively simple: those of reading the data written in the file, selecting and grouping them on command, and producing a pictorial display

of the result. In the terms used to describe manual processes in Chapter 3, they involve the techniques of reading data, grouping data (merge, dissolve, and generate), and performing data move and scale change on output. These techniques require a substantial degree of manual effort, and if a large number of cells have to be processed (as for the state-wide display of selected data items), the computer-aided processes would tend to be cheaper and quicker than manual techniques, even though inefficient in terms of computer usage.

The techniques of reading data, grouping data, and scale change referred to above, although high in manual motor effort, are nevertheless still in the low-order group (see Chapter 3, pp. 73-79) generally considered to be economically practical as working tools for a limited number of maps (such as might be required by a local planning agency). The system designers seem to have consciously or subconsciously recognized this fact and the file structure of LUNR is at its most efficient in retrieving data from small areas. The constraint on the system, however, is the low resolution inherent in the non-graphic data format, which seriously limits the utility of the system for analysis of such limited areas.

Data are added to the non-graphic store in the same manner as the original data. With small amounts of data this is an easy process, now that the files are established. With larger amounts of data, such as a new present land use survey, manual data extraction to create the non-graphic file would demand as much effort as for the original data set. Other than the ability to manipulate data as described above, the main overall benefit of the LUNR approach is perhaps the state-wide classifications of data now available in both graphic and non-graphic forms which have resulted from the data gathering activity underlying the creation of the actual system itself.

#### Arbitrary Regular-Area Systems - Type 2

The second type of system in the category of storage format using arbitrary regular areas is characterized by manipulative functions relying on a small cell size relative to the source data elements being stored. This type uses single data sets, or factors derived by a combination of such data sets, placed in small arbitrary grid cells which can then be spatially manipulated

by the computer. Such manipulation is both facilitated and constrained by the small grid format, and by the requirement for an end product with one data value per grid cell. The results are usually displayed as line-printer character maps.

SYMAP V<sup>1</sup> is the best known and most widely used program of this type. Other similar programs include MIADS<sup>2</sup> and MIADS2<sup>3</sup> (U. S. Forest Service), GRID<sup>4</sup> (Harvard University), LINMAP<sup>5</sup> and COLMAP<sup>6</sup> (Ministry of Housing and Local Government, U. K.), MANS<sup>7</sup> (University of Maryland), GEOMAP<sup>8</sup> (University of Waterloo), MAP01<sup>9</sup> (New York State, Department of Public Works), BRADMAP<sup>10</sup> (Bradford, U. K.), and numerous other unnamed, undocumented examples that exhibit the characteristics of the category and are in local use, for example the system developed and used in the Geography Department, University of Michigan, Ann Arbor, U. S. A.

Conceptually, the transformation from source data format to non-graphic format involves overlaying the source data with an arbitrary rectangular grid. The grid cells can be of any size, but they usually approximate to the area on the source map covered by one symbol on the line printer available. The only choice that can affect the resolution of the non-graphic store is exercised at this stage, and subsequent changes of scale only affect the area of paper used to display the recorded data set. The source data are then coded in the new graphic format of the arbitrary grid, the constraint being that only one value can be allocated to each grid cell. Points are assigned one cell each, lines are represented as strings of adjacent cells, and areas are described as clusters of cells. Further manual transformation to non-graphic format involves only the assignment of a simple coordinate system to the grid structure and the reading of the cell value for the appropriate cell coordinate. The resulting data set is thus metrically and topologically coded, with the relationships inherent in the array. The array may or may not be related to real-world location, that is, it may or may not have real-world dimensions which are known to the system. Manipulative characteristics of the system are usually independent of dimensions external to the array.

1. Synagraphic Computer Graphics, Laboratory for Computer Graphics and Spatial Analysis, Harvard University, U. S. A. SYMAP was originally developed by H. T. Fisher and Betty Benson, and subsequent modifications to the program have been made by A. Schmidt, R. Russell, D. Shepard, M. Manes and K. Reine. (See Harvard University 1968). 2. Amidon 1964. 3. Amidon 1966. 4. Sinton and Steinitz 1969. 5. Galts 1968. 6. Pearson 1972. 7. Pfaltz 1967c. 8. Steiner 1970. 9. New York State 1967. 10. Tomlinson, K. R. and Ituff 1972.

In practice,<sup>1</sup> a coordinate system of arbitrary plane grids is superimposed on the source map and the numerical codes of points, lines, or areas are recorded manually or by use of a mechanical coordinate reader. In some systems, areas (clusters of grid cells) can be coded by recording only their left-hand edge (assuming the map is entirely choroplethic), or by recording points which represent vertices of straight-sided polygonal areas. The process is slow and at best semi-automatic. The conversion of a base map of Montreal<sup>2</sup> (12 miles x 18 miles, 1:25,000 scale, 500 census tracts) to a non-graphic, small-grid format took 22 man-days. Similarly, the non-graphic coding of city blocks derived from 1:50,000 base maps for Paris<sup>3</sup> took 12 man-months.

Several types of image manipulation and data extraction can be carried out on the resulting non-graphic data. The implicit assumption in the majority of such manipulations is that the small grid cell size and single value per cell allow the cells to be regarded as an array of points without significant loss of information. This assumption is acceptable only if the cell size itself is small in relation to the real-world data elements. The boundary of the area around the point (the line segments forming the cell) are implicit, not explicit, within the system. The example that will be used to illustrate the practical problems of Type 2 of the category of storage format using arbitrary regular areas is the well-developed and widely applied SYMAP system.

#### SYMAP System - Example of Type 2 of the Arbitrary Regular-Area Category

In SYMAP V,<sup>1</sup> three fixed-format sequential files are established for each source map. The first file contains the coordinates of the cells that form the outer boundary of the mapped area, the second contains the coordinates of the cells that form internal boundaries within the mapped area, and the third contains the coordinates of cells that have data values associated with them. There are no explicit pointers in the files. Given these three files, the following types of basic manipulations can be carried out. Values provided at single grid coordinate points can be spread homogeneously over areas (clusters of grid cells) previously described to the system by the outer and inner boundary files. This process is akin to manual map colouring. However, if the values to be "coloured" are chosen so that unwanted values

1. Schmidt, A. 1966. 2. Seeman 1973. 3. Atelier Parisien d'Urbanisme 1972.

remain blank, then the process is equivalent to the manual data extraction procedure of reading, identifying, and retrieving specific data elements or data patterns.

For up to 1,000 points with data values attached, isolines can be calculated at equal intervals over the range of values of point data provided. The equivalent manual image manipulation process, contouring, is extremely efficient in terms of calculating the position of the isolines, as it involves a high degree of mental image processing; however, plotting the resulting lines is a relatively slow motor process. Comparable contouring of 500 points on a 20-in. x 30-in. map, for example, could be manually accomplished in 3 or 4 man-hours.

A third type of manipulation is location of the boundaries of spatial units, extracted from point information by nearest-neighbour methods. This is essentially a calculate-and-compare process carried out on the points in the data value file. The results are calculated for each row of the array and are printed sequentially. The manual equivalent is rather akin to a reverse form of centroid allocation, where the points are known but the boundaries between them have to be determined. If manual estimation is used, the mental selection process is quick, but the motor process of plotting the lines takes time. Furthermore, the error level in the estimation process is difficult to detect. If a more quantitative manual process is attempted, such as measurement between centroids to determine mid-positions, then extensive manual measurement is involved and the result still incorporates value judgements. Yet a third manual approach to the same task could be to emulate the computer process and lay a fine grid over the points concerned. The grid cells around each point could be coloured step by step until boundary conditions were determined. For any manual approach to such proximal image manipulation, the time to handle 100 points on a 20-in x 30-in. sheet would be in excess of 3 man-hours.

In contrast, the computer can manipulate an equivalent non-graphic store and produce all three types of data manipulation for a 20-in. x 30-in. map in approximately 50 sec<sup>1</sup> of IBM 370/155 central processing unit (cpu) time. Typical actual examples are retrieving and displaying all city blocks in Paris<sup>2</sup>

1. See Schmidt, A. 1972a for similar example. 2. Atelier Parisien d'Urbanisme 1972.



which costs between \$12 or \$75, depending on the type of manipulation (conformal, contour, or proximal). The generation of a SYMAP line-printer map of Canada<sup>1</sup> showing 258 census subdivisions, conformally shaded, costs approximately \$20. The production of a similar map for Montreal,<sup>2</sup> showing 500 census tracts, costs between \$2 and \$10.

Similar processing capabilities have been developed in other systems of the SYMAP type. A program called SECTION created at the University of Illinois<sup>3</sup> produces a cross-section profile along a row or a column of a SYMAP data base at the cost of approximately 20 sec of IBM 370/155 cpu time. A program named VIEWIT produced by the U. S. Forest Service<sup>4</sup> operates on a MIADS data base and calculates and displays the areas visible within a given radius from a given cell, based on elevation values assigned to the other cells. For an array of approximately 52 x 52 cells, this process costs 1 sec of UNIVAC 1108 cpu time.

These simple types of manipulation, which involve both image manipulation and data extraction from the non-graphic data storage, are thus efficient and inexpensive. They rely on the fact that the non-graphic data base is structured as an array, that the data can be handled as point values in an array, and that existing digital computers are particularly adept at array manipulation. Compared with manual techniques, the computers have advantages in their speed of calculation when any more than a few points are involved, and in their ability to plot the results of their calculations quickly on a line printer (up to 2,000 lines of characters per min).

Some SYMAP-type systems provide a very simple version of an overlay process by their ability to combine data values for one grid cell from different data sets (under the control of a set of rules) and to arrive at a new single value which is then recorded for the cell.

The use of SYMAP in conjunction with a program named SYMVU<sup>5</sup> produces an oblique-view map from the single data base. The maps are drawn by automatic plotting machines that use the output from the computer. Based on interpolation, these drawings show spatial variations in values of the data by variations in the height of a surface represented as if seen obliquely

1. Podelhl, 1973. 2. Seeman 1973. 3. Dudnik 1972. 4. Amidon 1972. 5. Harvard University 1968. 6. SYMVU is based on a program originally developed by Frank Rens. Subsequent modifications have been provided by Kathleen Reine, Lannon Leiman and other members of the Laboratory for Computer Graphics and Spatial Analysis (Schmidt, A. 1972b).

from above. The interpolated facts are shown without need for classing.<sup>1</sup> A further development of this capability for pictorial display is a program named OBLIX, by which contour lines are presented on the perspective view of the 3-dimensional surface. This latter capability, however, is only at the stage of active development.<sup>2</sup>

SYMAP V<sup>3</sup> has a wide number of statistical support options linked to the mapping system, which permit calculation of such analyses as means, standard deviations, histograms, and percentile groups of the data.

The earlier versions of SYMAP used larger-scale computers (for example, IBM 7094), but the current system can use medium-scale computers (such as the IBM 360/40 and limited versions can run on smaller machines.

The characteristics of SYMAP-type systems in this category are: relatively laborious preparation of the basic grid data; relatively coarse data representation (use of larger scales to achieve better resolution multiplies the manual preparation necessary); straightforward computational operations; simple but useful data manipulation capabilities on the very simple (one value per cell) data sets that they are able to handle; inability to handle complex data sets; and a simple graphic capability for display. The labour constraint on input usually limits the area handled by the system. The data are usually handled in punch-card form. Storage is not compact. The time required to respond to requests for manipulation and display of selected data sets is within a daily time frame, provided the base map is already coded. Computation times are low. Cost of data preparation is frequently four or five, and sometimes many more, times higher than subsequent computational cost of manipulation and display of the same data set.

The systems that belong to Type 2 of this category are useful for geographical tasks that involve few base map sheets and simple data sets; the manipulation and display of census data for urban areas is a typical application. Similarly, small-scale display of wider data distributions can be effectively handled. The systems are not efficient for storage and manipulation of extensive map series, and handling of line and network data and flow manipulations is difficult. The symbolic representation does not allow the handling of complicated descriptions, such as a forest soil capability with multiple

1. Fisher 1970. 2. Schmidt, A. 1970. Thomas, A. 1972. 3. Harvard University 1968.

subfactors. Even in situations where the capabilities of the systems can be put to their best use, application of the displays in a research or decision-making context is limited. In an academic sense, the techniques are valuable for raising questions about distributions, but are poor for providing data in a form on which decisions can be based. Possibly one of the most lasting benefits of these systems is their function of training their users in elementary techniques of non-graphic spatial data handling; another is in providing low-cost examples of computer-manipulated geographical data as an excellent visual aid, which serves to introduce the possibilities of further use of such capabilities in planning or administrative environments.

#### POINT DATA CATEGORY OF SYSTEMS

The second logical category of storage formats outlined in Fig. 4.7 is that of point data.

"Point data systems" handle real-world data elements stored in the form of irregularly distributed points. The inherent limitations of the types of operation that can be performed on the subsequent store of information were discussed earlier in this chapter (page 96). Nevertheless, the relative simplicity of storing and processing point data has led to use of the format as the basis for a number of current systems, including: FRIS<sup>1</sup> (Swedish Board for Real Estate Data); GIST<sup>2</sup> (Geographic Information System, New York); Coventry<sup>3</sup> (The Point Data System, Coventry, U.K.); NCC<sup>4</sup> (The National Capital Commission System, Ottawa, Canada); GRDSR<sup>5</sup> (Statistics Canada); and the proposed GISP<sup>6</sup> (Geographic Information System for Planning, U.K.).

As shown on Fig. 4.7, the use of the simple point format of storage is optimum for real-world point data but implies data reduction when used to represent areas (including clusters of points) and lines. To a great extent, current systems in this category accept the data reduction and use point data systems to manipulate area centroids as surrogates for data sets grouped by irregular areas. This is partly because relatively few of the types of real-world data elements that are of interest to investigators are truly point phenomena, and partly due to the complexity and expense inherent in higher-order systems that can handle lines and, particularly, irregular

1. Alfredsson et al. 1970, Melander et al. 1971, Selander 1970, Olsson and Selander 1971, Olsson 1971. 2. Amsterdam 1968, 1969. 3. Coventry City Council 1970. 4. Symons 1972a. 5. Statistics Canada 1972a, 1972b. 6. Joint Local Authority 1972.

areas. It remains to be seen whether the system users can accept the information reduction inherent in the point storage of higher-order data formats.

The simplicity of point format processing has led to its early adoption by agencies wishing to establish some form of location identifier for large data sets (for example, 20 million records, 120 data values per record<sup>1</sup>). The result has been the development of relatively sophisticated procedures to aid the process of allocating variously coded data to the appropriate points. Even though the initial stage of creating the non-graphic data file usually relies on manual or semi-automatic techniques, the most significant contribution of the current systems lies in the subsequent file creation processes. They have, in fact, made it possible for location identifiers to be assigned to data sets that could not, for all practical purposes, have been displayed in cartographic form by conventional manual techniques (due to reasons of time, human resources, and expense, rather than to technical capability). There is, as can be understood, some difficulty in comparing the computer-aided techniques with manual-aided techniques without assuming a graphic data base that could not be created economically by manual methods, though some comparisons will be attempted later in this discussion.

Systems using the point data approach commonly use a tracing-type digitizer to provide a coded metric representation of existing cadastral maps. The techniques used involve the initial creation of a set of coordinate points that can be used as index points for the data sets or discrete data. A medium-to large-scale computer is most often used to store and manipulate the data. Various approaches can then be used to relate data to the coordinate points. Retrieval of data is usually by arbitrary polygon or network, and display of data is usually accomplished on a line printer. Because of the large quantities of data manipulated in some systems, efficient file structures have been developed for the selection and retrieval of spatial data from large files. The time for response to requests can be rapid enough to handle day-to-day queries involving data summary by area. Requests repeated frequently are amenable to a standardized query format. These general system characteristics have led to the adoption of point storage systems for handling quite complex and closely spaced data sets from relatively small areas. They are

1. Statistics Canada 1972a, 1972b.

commonly developed to handle urban data and can be used in small-scale representations of rural data.

One system that illustrates the main characteristics current in the category of storage format using point data is GRDSR, the geographically referenced data storage and retrieval system of Statistics Canada.<sup>1</sup> It is briefly described below.

#### GRDSR System - Example of Point Data Category

The overall objective of GRDSR is to facilitate the retrieval of subsets of census data by small areas. The census (Canada 1971) comprises approximately 21 million small data sets (120 data values per set of data), topologically coded (by postal address, and enumeration district and area numbers) to an external graphic index. In urban areas, street addresses can be related to a street map; in rural areas, enumeration numbers can be related to enumeration area maps. Both street maps and enumeration area maps are based on UTM-coded cadastral sheets.

The initial step in conversion to non-graphic format is coding the external graphic index topologically and metrically. Within urban areas, the street network is topologically coded by street name, node number, and civic address range of street segments, and is metrically coded by node coordinates. The coordinates of points to which data from one city block side (face) will eventually be allocated are internally calculated from the metric node codes, according to a given set of rules. The resulting non-graphic definition of the city is termed an area master file (AMF). Thereafter, the street addresses that identify the data sets are internally standardized by a program called the postal address analysis system (PAAS), matched to the topological codes of the AMF (through the address conversion file), and, if the match is successful, are assigned the related metric code (the block face point coordinate). Verification and editing procedures are built into these steps. In rural areas, centroids of population density are manually assigned to each enumeration area (this being an extremely efficient manual process, as noted in Chapter 3) and the metric coordinates and enumeration area codes of the centroids are recorded semi-automatically. Records from rural areas are assigned to the metric coordinate via the

1. Ion 1969, Statistics Canada 1972a, 1972b.

topological code of the enumeration area number. Deriving the non-graphic topological codes and metric codes for both urban and rural areas is a slow, semi-automatic process, which is characteristic of the category of point storage systems. The FRIS system, for example, which uses an approach similar to GRDSR for assignment to rural area coordinates, but carries it to the level of land ownership parcels, has required six two-man teams of digitizers to cope with the data at a rate of approximately 30,000 parcels per year, and it is estimated that there are approximately 3 million land parcels in the populated parts of Sweden alone that require further point allocation.<sup>1</sup> In Canada between 1968 and 1972, 14 urban centres were partially coded, 225,000 block faces were identified, and 27,000 rural enumeration areas were coded. No measures of the effort required to accomplish this task are available. The assignment of records to the derived point coordinate codes, however, is efficient. The total cost of address standardization, verification, and coordinate assignment in GRDSR is approximately 1 cent per address.<sup>2</sup>

The file structure<sup>3</sup> of GRDSR is straightforward, but illustrates how the efficiency of subsequent data retrieval is dependent on an appropriate file structure. The declared objective of the GRDSR system is to facilitate the retrieval of subsets of census data (or other data) by small areas. Inverted data grouping of the basic data set produces 120 named, indexed, sequential data strings (subsets) of mixed-format data values (fixed format within each string). Each data string is stored on a direct-access storage device with its appropriate control information (header, level,<sup>4</sup> format, length, and so forth). The internal index (called the data directory) lists the block face point coordinates and uses intra-data-element pointers that relate the coordinate points to individual data values. One data directory is created for each city. The advantage of the data plus pointer index (data directory) is that only the required strings are accessed for each retrieval request. Within the string, only the required elements are actually read. The data directory itself is indexed by the query area library (QAL), which consists of named sets of pointers derived from the data directory. The names of the pointer sets in the QAL hence define specified areas for data retrieval and can be used as statements in the retrieval command language (TARELA).

1. Cripps 1970. 2. Statistics Canada 1972b. 3. Podehl 1972a. 4. See p. 113 for explanation of "level".

Retrieval<sup>1</sup> is limited to gathering specific data values for specified points. No other form of data manipulation in the spatial sense is possible. Areas are specified for retrieval by any named set of pointers in the QAL, or by a set of pointers generated by matching the data directory with a metrically or topologically defined retrieval area by a point-in-polygon routine (PIPA). If the retrieval area is topologically defined, by street names or node numbers, they are matched against the AMF to produce coordinate points which are used to define the polygon used in the PIPA. Formulas of geometric shapes can similarly be used to define a retrieval polygon. Each set of pointers generated by such requests is named and added to a temporary or permanent form of the QAL; thus a "library" of frequently used retrieval areas is on hand and previously specified areas need not be respecified.

Data values are named in a data dictionary, and the names can be used as statements in TARELA.<sup>2</sup> As the data strings are to some extent hierarchical (some contain data relating to households, some to persons, and so forth), their "level" is coded in the string file and the file system (STATPAK<sup>3</sup>) allows the internal generation of inter-file pointers to permit topologically directed search (for example, for values attached to specific persons within certain households). This again improves the search efficiency of the system. Output from most requests is generated in the form of tables.

Values for specified coordinate points can be fed (via a program called MAPPAK) to SYMAP for display. The AMF can be automatically plotted by a program named MAPMAKER.

Overhead costs are involved in extracting area definitions from QAL, in analysing the TARELA request, and in the compilation and linking-editing phases. The execution costs are closely dependent on the number of values retrieved from the strings. Typical tabulations range from \$30 to over \$100.

The contributions of systems in the category illustrated by GRDSR rest in their ability to assign large data sets to location identifiers and to select and retrieve such data efficiently. The spatial manipulation capability currently incorporated in the system is trivial, even taking into account the inherent limitations of manipulating point data. An additional consideration is the reduction of data involved in assigning groups of point data (areas) to single

1. Statistics Canada 1972c. 2. Phillips, J. 1972a. 3. Podehl 1972b.

points. While this facilitates storage, as can be seen above, it imposes substantial constraints on the use of data. The well-known point-in-polygon retrieval process inherently produces spurious results when points are surrogates for unknown, irregular areas, unless sufficient points are included to minimize the edge effect. In a system whose objective is to produce small-area data, the point-data format is thus the major constraint on reliable data retrieval. The acceptance of this fact may be related to the requirements of confidentiality of census data (even though in the GRDSR system there is an additional inbuilt process, not described above, to ensure such confidentiality<sup>1</sup>). Nevertheless, the ability to retrieve data from large files in not-so-small areas specified by the user is a great step forward from the state of availability of census data only by predetermined aggregations of considerably larger size, providing the limitations of the approach are understood.

As mentioned above, the GRDSR system uses points as surrogates for areas (clusters of point data), which are undefined in the retrieval system, which can be of widely differing shapes and sizes, and which may or may not be externally defined (for example, mapped). Other systems in this category, such as FRIS and NCC, measurably increase the effort needed initially to code the data but utilize points as surrogates for smaller areas (land parcels) which are externally defined and in some instances are known to the system. Some additional spatial data manipulations are also possible in other systems (for example, measuring the areas represented by the dots in FRIS<sup>2</sup>), but they are generally limited.

Comparison of the capabilities of a system in the category exemplified by GRDSR with the capabilities of manual graphic data handling techniques poses the problem of selecting a starting point for the comparison. The true basic data sets are, of course, the handwritten census returns from individual respondents. Prior to 1968, the only form of graphic display of census data in Canada was individually produced maps of census data shown as values within census subdivisions at a considerably higher level of aggregation than enumeration areas or city blocks.

With the advent of large-scale computers in the 1960s, the basic census data were transcribed in computer-readable form to aid the tabulation process. It

1. Phillips, J. 1972b, Murphy 1972. 2. Olsson 1971.



is the availability of that large amount of data in computer-readable form that makes the design of GRDSR practical, but whether the transcription costs should be assigned to the system is debatable. The graphic equivalent to the non-graphic store of data as coded for GRDSR would be a series of maps with the point locations plotted and the data values for each point indicated (either as clusters of values on sheets or on separate sheets, 120 in all). Given such a graphic file, the manual techniques that parallel those available in GRDSR would be the superimposition of polygons, and the counting of data values within those polygons. The manual operation would be quicker and cheaper than GRDSR if very few data points were required from very few sheets, but would rapidly become more expensive as the search area or areas multiplied and the number of data values increased.

The image manipulation capabilities inherent in SYMAP are part of the display possibilities of GRDSR. The relationship of those techniques to manual processes was discussed in the preceding section of this chapter and applies here. In general, the contribution of systems in the point data category to current spatial data handling requirements lies in their topological sophistication rather than their metric facility.

#### LINE SEGMENT CATEGORY OF SYSTEMS

The third category of storage format contains line segment data systems, which are those with the capability of handling real-world data elements stored in the form of line segments. The boundary conditions separating line segment data systems from lower-order systems are that the line segment networks must be defined in the primary file structure of the systems and that data so related can be manipulated and extracted without further network specification. Although it is convenient to establish this subdivision in the discussion of current systems, the limitations of the categorization should be noted. First, the higher-order systems, such as line data systems, are not necessarily unimodal. They may incorporate in their design the ability to handle data in formats other than their primary format and to change one format to another. (The difference between those two functions should be recognized.) Secondly, as was mentioned at the beginning of this chapter, the development of techniques to handle spatial data

is in a state of transition. Systems originally designed to be unimodal are being changed to handle additional data formats. An example of this latter type of development is the FIUS system, which was originally designed as a point data system to link administrative data files and which is now having added to it the capability of handling straight line segments in networks. In the interim, the general grouping of systems into categories by primary data format will be retained for convenience of description rather than as a proposal for formal ordering.

Several current systems were designed with line segments as a primary data storage format. Typical examples are: SACS<sup>1</sup> (Street Address Conversion System, Urban Data Center, Seattle); Vancouver System<sup>2</sup> (City of Vancouver Engineering Department); AUTOMAP<sup>3</sup> (Central Intelligence Agency, U.S. Government); and DIME<sup>4</sup> (Dual Independent Map Encoding System, U.S. Bureau of Census). Two systems will be used to illustrate the range of characteristics within the category of storage formats using line segments. The DIME system is typical of the category in that it attempts to handle substantial volumes of data related to each line segment. The line segments are thus used to form a spatial structure to which other data can be related. The second example is the AUTOMAP system, which is typical of systems developed to provide a capability for manipulating the images formed by the line segments. The emphasis of these two types of system is different, but both use line segment networks as their primary file structure. Both are well-developed systems and are in daily use for purposes other than those of system development. The first to be examined is the DIME system.

#### DIME System - Example of Line Segment Category

The DIME system views a map as a combination of linear graphics and alpha-numeric data.<sup>5</sup> The linear graphics can be described by points at which the rate of change of curvature of the line exceeds some given bounds. These points characterize in abstract form the structure of the graph, a structure determined entirely by implied metric-free relationships. DIME topological coding begins with division of the area to be coded into an exhaustive and mutually exclusive set of bounded regions. Each of these regions is assigned a unique nominal code. Regions may have single or multiple boundaries.

1. Calkins 1967, Barb 1968. 2. City of Vancouver 1965. 3. Schmidt, W. 1969.  
4. U. S. Bureau of Census 1972. 5. Corbett and Farnsworth 1971.

This division requires that an area to be coded be treated as a single, simply connected surface. The boundaries between regions then form the minimum line (segments) network which must be coded. Any additional lines (interior to a single region) may also be coded as the application requires.

All points (nodes) where lines meet are then assigned unique nominal codes. The minimum coding for each segment then consists of four nominal codes: one for each end point and one for each side. Since no segment may cross another, a side of a segment is entirely in one and only one region.

The sequential arrangement of the two node codes within the line segment record is used to specify orientation. A segment is considered to be oriented away from the first and toward the second node. This orientation may be entirely arbitrary, but once set down it determines the order of the region codes, the first of which is that of the left side of the segment, the second the right side. Normally the region boundary lines are streets and therefore it is useful to record street names and address ranges as additional nominal codes, and to break up the boundaries by street names. A wide variety of other nominal codes can be assigned and related to the nodes, the segments, or the sides (areas). The resulting individual line segments are in fixed format. The connectivities that are specified depend on the selection of nominal codes which are recorded as part of each record, but in essence, two networks are described. One network is that of the segments' connecting points, the other that of the segments' connecting (separating) regions. The principle of duality leads to some efficient internal editing checks. The main edit routine of the DIME system is the program (KIRCHOFF) that uses the Kirchoff algorithm for determining the connectivity of the graph. After assembling all segments with a particular region code (on either side), one must be able to form one or more closed chains with segments chained together by the node code. If any bounding segment has been omitted, or if any code has been recorded incorrectly, or if the orientation is incorrect on any segment, then the chaining will fail. Similarly, one can assemble all segments with a given node code and try to construct a chain around the node with the segments chained on region codes. Nodes must be singly bounded if there are no duplicate codes.

The file can be metrically coded by adding the coordinates of nodes to the segment records. Coordinates are added to the files only after a high level of internal topological consistency has been achieved.

The DIME system has been developed by the U. S. Bureau of the Census. It has two significant advantages that are particularly valuable, one of which arises from the topological nature of its coding and the other from its census origin. The size of the United States has inhibited the development of a uniform national grid that would support a map file encoded primarily by tracing. Any information system that is created on a geographical basis at present has to reckon with source data of a variable and uncertain quality. Hence, there are strong advantages inherent in a spatial data system which can exploit topological relationships to the fullest extent.

The DIME coded data set is amenable to various types of file structure, the choice of which depends upon the particular application. This is illustrated by three examples below. The U. S. Bureau of Census basically uses a postal census for its urban areas (at least 60% of all households in the United States respond in that manner.) This led to the early development of listings in cities of postal addresses that were grouped by block faces. Each block face file in this address list (ACG, address coding guide)<sup>1</sup> is identified by state, county, census district, city, zip number, street segment number, street address building numbers (at each end), administrative code, and an order number. ACG files have been established for 225 cities in the U. S. A. The line segment DIME records can be sequentially arranged by node numbers within street names. This step produces two lists of block numbers, one for each side of the street. The census records (address coding guide), coded by street address and block codes, can then be matched into the segment records by matching street names and block codes, using ADMATCH or UNIMATCH programs.<sup>2</sup> Hence, all individual census records can be spatially assigned to line segments within the city concerned.

The DIME approach allows collation of separately established external indexes such as city directories<sup>3</sup> with the DIME file itself, on the basis of street names. A city directory file consists of an address coding of contiguous segments within street name with identification of intersection streets. The DIME data set is sequentially arranged by node numbers within street

1. U. S. Bureau of Census 1971a. 2. Jaro 1972a, 1972b. 3. U. S. Bureau of Census 1971b.

names, as described in the previous example. The resulting chain of node numbers is used as an index to retrieve the names of intersecting streets from the original file and add them to the new list. The resulting file is in essentially the same structure as the city directory, and it is possible to collate the two files on the basis of street segments between intersections.

The third example concerns the collation of two DIME files for the same area, which have been coded for different purposes and contain different codes. It is assumed that the two encodings are independent as to node numbering, and that the area codes recorded are different in each case (the twin cities problem). The collation proceeds by assembling the street names at each node independently for each file, and then comparing the groups of names to develop correspondence between node numbers in both files. Even if both files contain coordinates, the topological approach for developing correspondence is probably a more reliable procedure than matching coordinates, given the inevitable differences in coordinate readings from two independent maps.

The DIME system has its disadvantages. It is of very limited usefulness in rural areas, as the rural road network leaves interstitial areas poorly differentiated and rural postal addresses are very difficult to allocate to specific line (road) segments. The ACG does not exist in rural areas. There is considerable difficulty in relating the DIME file to certain kinds of maps, such as contour maps, which can only be collated on a coordinate basis.<sup>1</sup>

The cost of creating a DIME file, including the coordinate file, is approximately 50 to 70 cents per block. This includes the updating and correction of the available maps, coding of the nodes, allocation of street segment descriptions, card punching and transfer to magnetic tape, a validity check on the file, and the addition of coordinates. It usually requires a large-scale computer to operate efficiently. However, the edits will run on an IBM 360/40 with as little as 32K bytes of storage.

Undoubtedly a major benefit of an established file in the category of systems like DIME is the subsequent ability to collate spatially the data sets of interest to an investigator. An excellent example of this is the DIME-based health information system<sup>2</sup> (HIS) established in New Haven, Conn., U.S.A. Nearly

1. Observatoire Economique Méditerranéen 1971. 2. U.S. Bureau of Census 1970a, 1971c.

300 sets of data items from such sources as the Census First Count Summary Tapes, City Vital Records, and the Mental Retardation Register were linked through ADMATCH and summarized by block group. Correlational, factor, and further multivariate analyses were performed on the data to produce a few constructs or typologies.<sup>1</sup> The resulting typologies themselves and maps displaying which block groups are typical for a given typology (that is, which block groups rank in the top quartile among all the block groups with respect to this typology) were produced for city health service planning. A generally equivalent manual process could be considered to be the manual allocation of the selected data values identified by street addresses to city blocks (or subjectively defined city zones) with subsequent manual counting and summary leading to the equivalent statistical analysis, followed by generating appropriate maps of the results. An alternative might be to plot each data set as points on a city map, with subsequent visual interpretation of point cluster characteristics. Either approach would involve extensive manual effort and it can be noted that despite the long-term availability of the statistics concerned, no attempt has been made to carry out this type of spatial analysis by manual methods (though such a decision is influenced by the level of city management processes as well as the cost of cartographic techniques).

The most sophisticated spatial analysis techniques based on the networks defined in a DIME system are incorporated in a program known as the computerized resource allocation model (CRAM). CRAM<sup>2</sup> is a generalized system for determining the service areas for a set of facility locations. In its most general form, service areas are constrained by facility capacities and travel times. Given data on demand and facility capacity, the problems that can be attacked through CRAM run from simple fixed-source districting problems, such as allocation of school district boundaries, park planning, and site location, to much more complicated problems of emergency vehicle routing, delivery or bus route planning, and freeway location studies. CRAM uses a modified version of the Moore shortest path algorithm<sup>3</sup> to do its geographic analysis; the modification consists of changing Moore's technique of finding the shortest path between points to one of viewing the network as a

1. A typology in this sense is a synthesis of many items, ordered logically by their contribution to the whole. 2. U. S. Bureau of Census 1971d.  
3. Farnsworth 1970.

set of interconnected links, the connections being nodes. This approach fits the algorithm to the primary DIME file structure and brings some advantages in the allocation of demand, because disaggregation of persons (demand) to links is more reliable than disaggregation to nodes, as links have length and the disaggregation may be varied according to length, but this cannot be done with nodes. In manual terms, the use of graphic storage as a source of data for such spatial analysis would imply the manual measurement of length of all relevant street segments on a street map, the manual allocation of population data from individual census (or other) records by street address to the line segments, and the identification and calculation of shortest paths in the simple case of a fixed facility. It is doubtful if the iterative processes of optimizing collection routes or allocation of multiple facility locations would be attempted quantitatively by manual methods. It is much more probable that estimation, coupled with a limited amount of manual measurement, would be employed to avoid an incalculable amount of manual effort. In either of these manual approaches to network analysis, a high percentage of the manual tasks are of a high order because the data required are inherent in the graphic file structure, rather than being explicitly written in it.

More direct comparison between manual techniques and computer-aided techniques related to the DIME type can be made from an example of the simpler spatial data manipulation tasks. Using the DACS program<sup>1</sup> (DIME Areas-Centroid System), the size of city blocks, census tracts, or other areas nominally defined in a metrically coded DIME file can be calculated. This would compare with measurement of the same areas by dot grid or planimeter on the graphic version of the map. Similarly, by the DACS program, the position of area centroids can be allocated and their coordinates calculated. This would compare with the efficient process of visual area centroid allocation coupled with the relatively slow manual process of determining coordinate values from an overlaid grid. Typical figures for the combined computer-aided area calculation and centroid allocation process (producing a listing of area identifier, size, and centroid coordinates) are: for 288 blocks in Akron, Ohio,<sup>2</sup> 1.6 min of IBM 360/40 cpu time at a

1. Meyer 1972. 2. White, M. S. 1973.

total cost of \$4.00; and similarly for approximately 5,000 blocks in New Haven, Conn.,<sup>1</sup> a total cost of approximately \$80.00. (Costs are not quite linear as the process involves both sorting and calculation, and the cost of the sorting rises faster than the cost of the calculation.) It is also possible, thereafter, using DIME system collation procedures, to allocate the data summarized by blocks or other areas to such calculated centroids. Point-in-polygon retrieval can be carried out subsequently on such point data (in exactly the same manner as the GRDSR point data system described above, and with all the constraints inherent in point data systems applying to the process). Even assuming a convenient way of manually allocating data sets to graphically stored points, the manual process of such polygon overlay and point counting requires considerable motor effort if the area concerned is large or if the process has to be repeated for various types of polygon.

As a complement to the DIME type of system, which uses a line network as a basis for manipulating related sets of data, it is necessary to consider the type of system that incorporates the ability to perform spatial data manipulations on a relatively simple image file of line segments and points but that does not provide multiple linkages to related data sets. A typical example of this part of the line segment category of systems is the AUTOMAP<sup>2</sup> system of the Central Intelligence Agency, U. S. A.

#### AUTOMAP System - Example of Line Segment Category

The AUTOMAP system concerns itself with the extraction of simple line and point data from existing maps, with their storage in a non-graphic form, and with the ability to select and manipulate the set or a subset of the non-graphic images (image manipulation techniques) in quite a comprehensive way before displaying them again in graphic form. It does not, as far as is published, have any extensive ability to extract data from the non-graphic storage if they are not already written in the file (see data extraction techniques).

The graphic source data originally used were the U. S. Navy Oceanographic 12 Sheet Map "The World" (1:12,233,000 scale) and the U. S. Air Force Planning Chart "Antarctica" ASC-6 (1:9,000,000 scale). The data elements that were non-graphically coded were line segments describing international

1. White, M. S. 1973. 2. Schmidt, W. 1969.



boundaries, coastlines, rivers, coordinate grids, and selected point data.

Point data elements are metrically coded by latitude and longitude coordinates and are nominally coded by name, type of location, "rank", map area code, and accession number.

Line data are metrically coded and transferred from graphic storage to non-graphic storage by using a digital line follower, a Thompson PF10LDS, which creates a string of arbitrary X,Y coordinates for each line segment. Each line segment is further identified by a number uniquely assigned within each continent. The line segments are placed in a variable-format indexed sequential file structure in order of segment number. The index relates the line segment number to a map area code, a feature type code, and a "rank" code.

The "ranking" code allocates each data element to one of a hierarchical series of subsets of data. The codes are thus intra-file pointers to an external directory of rank values. The ranking code coupled with the names of the types of data element required are the necessary data descriptor statements in subsequent retrieval and manipulation processes.

A mathematical transformation is subsequently performed on the line segment file to convert the arbitrary coordinates to geographic coordinates and radian values. The geographical coordinates of the point data sets can similarly be transformed to radians. This, then, is the first example in the set of systems described in this chapter that ventures into any form of three-dimensional metric coding. The system can carry out these input transformations on source data arranged in plane graphic storage in any one of four common map projections.

Once the spatial data are organized in this manner in the non-graphic file, it is possible to carry out image manipulations that are both interesting and extensive. Using the coordinate values and radian values, the non-graphic store can be rearranged into any one of 16 plane transformations by the CAM<sup>1</sup> (Cartography Automatic Mapping) program. These projections are:

1. Schmidt, W. 1970.

1. Mercator - sphere or spheroid
2. Miller Cylindrical - sphere or spheroid
3. Transverse Mercator - sphere
4. Original Input Projection, regardless of type (one-to-one)
5. Azimuthal Equidistant, Oblique - sphere
6. Orthographic from Infinity
7. Equirectangular
8. Stereographic - sphere or spheroid
9. Lambert Conformal Conic with Two Standard Parallels - sphere or spheroid
10. Ptolemy Conic Equal Interval with One Standard Parallel
11. Kavraysky IV Conic Equal Interval with Two Standard Parallels
12. Alber's Conic Equal Area with Two Standard Parallels
13. Transverse Mercator Spheroid
14. Gnomonic - sphere
15. Perspective from a Given Altitude
16. Azimuthal Equal Area

It will be remembered that the manual process of changing map projections was extremely lengthy and laborious and was a function usually undertaken only by large cartographic institutions. The AUTOMAP system can carry out such tasks within a single day.

A further interesting ability is to select specific areas for retrieval and at the same time to specify topologically the categories of data that are required for display, and also, if necessary, incorporate a generalization of the linear data. As mentioned above, the selection of items to be displayed is based on the codes for data type and rank level, such as all towns (type) of over 100,000 inhabitants (rank). Linear generalization is achieved by ignoring segment coordinates at specified intervals and assigning a free-run (interpolated) plotter line through the remaining points. This approach to linear generalization is simplistic but adequate for the decision-making purposes that the system was designed to accommodate. (The coastal island problem is dealt with by a simple algorithm.) Dependent on the requirements of the user, the system has the ability automatically to search the file, to carry out the calculations necessary for the required projection, scale, and output device, and to filter to eliminate superfluous points. This process thus not only includes the image manipulations of the data base, but incorporates selective extraction akin to the process of tracing selected data from graphic storage.

The system has the further ability to generate and draw projection grids, range rings, azimuths, and circles or spirals around a given point, and to superimpose these on the data to be displayed. In the same manner, it can generate and draw rectangles, map boundaries, great circle paths, ellipses, a variety of special symbols, map centre points, corner ticks, and title blocks.

The system employs a large-scale computer (IBM 360/65) and can produce graphic output suitable for a variety of output devices: line printer, line plotter, and cathode ray tube. To do this, it can control a light-beam drawing head on the plotting table, control variously coloured pens on one output, and control the size and slant at which special symbols are drawn. The time between request for data and a final map plotted to the specifications requested can take as little as 3 to 4 hours. The system has an extremely limited suite of data sets with a total number of points and line segment nodes in the order of 100,000; it cannot handle the printing of names with any facility, and has no capability for measurement or comparison of data.

The system is by no means unique. Similar approaches have been used at several agencies, including the Experimental Cartography Unit,<sup>1</sup> R. C. A., London, Rome Air Development Center,<sup>2</sup> New York, the Canadian Hydrographic Service,<sup>3</sup> Ottawa, the Geography Department, University of Michigan,<sup>4</sup> the Ontario Department of Highways,<sup>5</sup> Toronto, and the Electromagnetic Compatibility Analysis Center,<sup>6</sup> Annapolis, Maryland. As the use of such map compilation systems increases, so there will be a build-up of cartographic information in digital form that will be amenable to spatial analysis of many kinds. Although this type of analysis is not the aim of many systems of this category at present, it may be a future benefit which far exceeds the current objectives of economy and facility in producing cartographic images. Some limited measurement and comparison of non-graphic storages are being carried out in some related systems. The digital information is, for example, being used to calculate distances, areas, and volumes in various applications. Profiling,<sup>7</sup> line-of-sight analysis,<sup>8</sup> and radar-masking analysis<sup>6</sup> are also being carried out. The AUTOMAP system

1. Bickmore 1971, Lang 1971, Rhind 1971. 2. Diello et al. 1969, Diello 1970. 3. Boyle 1970b, 1972. 4. Tobler 1970. 5. McLeod 1972. 6. Electromagnetic Compatibility Analysis Center 1969. 7. Clark 1970. 8. Boehne 1967.

does, however, represent one such system that is in day-to-day use for purposes other than its own development.

#### SUMMARY

This chapter has focussed on the limitations and capabilities of current techniques for non-graphic storage and computer-aided manipulation of spatial data. The basic attributes of non-graphic storage were briefly discussed. General categories of current computer-aided techniques were then examined and their characteristics illustrated with specific examples. Computer-aided techniques were compared with manual techniques. Several conclusions can be drawn from the evidence presented.

Assuming that the large-scale computers common in most research and government institutions in the 1970s are available to the user, then the factor that determines the usefulness of current systems is the format and file structure of the spatial data stored in them. The primary shortcoming of current systems is the constraint on development of the format imposed by the economics of transforming graphic source data to a metrically coded, non-graphic form. In all systems described the process is manual or, at best, semi-automatic. In typical lower-order systems, extremely high levels of manual effort are used to create data formats with inherent information loss, and file structures of the simplest kind. In the higher-order systems the creation of the basic metrically coded storage formats involves effort ranging from 5 to 50 man-days per 20-in. x 30-in. map sheet. The effort involved puts severe constraints on the amount of data available to the computer. It should be emphasized that this applies primarily to the metric aspect of non-graphic coding. The most significant advance in overcoming the limitations of current practices for converting graphic to non-graphic data lies in the ability to relate spatial to topologically coded data sets, developed in the higher-order systems exemplified by PAAS,<sup>1</sup> ADMATCH,<sup>2</sup> and UNIMATCH.<sup>3</sup> The limitation on metric input is significant as it can have the effect, at present, of reducing the resolution of the input data (in an attempt to reduce the input effort) and thus imposing significant control over the subsequent utility of the computer-aided process. Similarly, it limits the volume of data available to the system, and the

1. Statistics Canada 1972b. 2. U. S. Bureau of Census as described by Jaro 1972a. 3. U. S. Bureau of Census as described by Jaro 1972b.

primary advantage of computer processing is its ability to handle rapidly large volumes of data. Certainly, if improvements in current systems are to be expected, they would be of greatest benefit in the input process.

In contrast to the limitations imposed by the transformation of data to non-graphic formats, there is a significant under-use of capacity in subsequent data manipulation and extraction. In all categories of system so far described, the file structures are at best straightforward. It appears that the capacity of current (1974) large-scale computers to manipulate data is considerably in excess of the manipulations x volume of spatial data that they are being asked to handle. This may well be a constraint imposed by the input limitations described above. It may be a lack of requirement for data to be manipulated in a sophisticated manner (that is, a lack of user need or competence). In any event, there is no evidence from the file structures examined so far that they are designed to facilitate (that is, to make possible economically) complex spatial data manipulations.

These considerations lead to the general observation that current computer-aided systems are, to a great extent, conceptually extensions of manual processes. As mentioned at the beginning of this chapter, they are in a transition phase between manual techniques and considerably more advanced computer-aided processes, the latter using either coded or non-coded information. At the end of the chapter concerning manual techniques, it was noted that increasing data volume rapidly decreases the type of manual technique that can be considered economical, and it was suggested that computer-aided techniques might profitably seek to reduce manual motor processes and allow the presentation of data to a user in a form that does not overload his channel capacity (the last two suggestions being a direct function of the volume of data to be handled). The key to the utility of computer-aided systems is thus the degree to which they can spatially manipulate increasing volumes of data. Such spatial manipulation can take two forms. The first of these is the ability to reduce the information content of images, or reformat them to fit mental indexes, to the point where their visual examination does not exceed human channel capacity (so that the maps produced become clear or are comprehensible to mankind, but

where the subsequent correlations or patterns of spatial data are still visually determined by the observer). The second type of spatial manipulation is the ability to retain the total information content of the images and systematically analyse them non-visually (because the volume of graphic data does not allow visual inspection), then to present the results of such analysis to the user.

The lower order of current systems and the higher order of image manipulation systems (such as SYMVU,<sup>1</sup> OBLIX,<sup>2</sup> AUTOMAP<sup>3</sup>) make their greatest contribution in the first mode of spatial manipulation. That is, they have as their objective the display of medium to large data sets for subsequent visual synthesis. They bring somewhat larger data sets than can be normally visually accommodated to a point where human channel capacity is not exceeded. In this process, they can accept, in fact must incorporate, some information loss. This type of system, despite the use of computers, is considered to be conceptually manual (visual).

The higher-order categories of current systems approach the point where the data they currently contain cannot be visualized economically and where the second mode of spatial manipulation must be employed. At this point, by reason of volume of data (a product of data resolution, size of data sets, and number of manipulations required), it is literally impossible to conceive mentally the result. (The primary factor is one of human channel capacity.) It is considered that the higher order of current systems is just entering this area of capability. Some systems have large data files but are unable to manipulate them other than to retrieve subsets of data efficiently. Other systems, however, such as the CRAM program operating on a DIME file, do approach this level of capability. One can perhaps look forward to the time when instead of carrying out an optimizing analysis on a plane network in two dimensions it will be possible to have similar analysis of multilayer networks, or three-dimensional networks that generally define volumes in space, or dynamic multidimensional networks. In addition, or as an interim step, one might anticipate systems that can incorporate both models of spatial manipulation on an interactive basis. The current systems are truly in an

elemental state, but give some indications of the capabilities that can be expected.

The following chapter will examine somewhat more advanced systems that are now being developed.

## CHAPTER 5 - COMPUTER-AIDED SYSTEMS UNDER DEVELOPMENT FOR HANDLING MAPPED AND OTHER LOCATION- SPECIFIC DATA

Computer-aided techniques for handling mapped and other location-specific data are described here as being "under development" if they have not been implemented as data handling systems in agencies other than those where the techniques were developed. They are generally regarded as being still in the development stage by the staff concerned with their design, and usually are not documented to the level where transfer of the technology to other agencies could be easily accomplished. This does not mean that useful interim products are not generated by such systems in their current state of development.

The term "under development" necessarily covers a large and heterogeneous realm of activity and there are difficulties in providing an integrated view. The first half of this chapter will continue the discussion of storage format categories from Chapter 4, examining the current work on systems with storage formats that use irregular polygons; the Polygon Information Overlay System (PIOS)<sup>1</sup> and the Canada Geographic Information System (CGIS)<sup>2</sup> will be taken as examples to illustrate the category. The second half of the chapter will examine more general current developments in input (digitization), manipulation, and output capabilities which apply to all the categories of system discussed so far. The chapter will conclude with a discussion of the modern trend toward multi-format systems which combine the properties of other systems described here. In Chapter 6, the "state-of-the-art" will be examined by reference to the progress represented by all of the developments discussed in Chapters 4 and 5.

### IRREGULAR POLYGON CATEGORY OF SYSTEMS

Systems that handle real-world data elements with irregular polygons as their primary storage format are generally considered to be still (1974) under development.

The boundary conditions that distinguish irregular polygon data systems from lower-order systems are that the irregular polygons must be separately

1. Dangormond 1971. 2. Tomlinson, R. F. 1967.



defined in the primary file structure, and that the data related to them can be manipulated and extracted without further polygon specification. Such systems are not necessarily unimodal. They may incorporate in their design the ability to handle data in formats other than the primary one, and to change from one format to another.

Irregular polygon data systems are characteristically used to store and describe the areal attributes of a surface. The line defining a polygon boundary has regions on either side of it with different surface attributes, whereas lines defining point connectivities (networks) may have the same attributes on either side (that is, the network may pass through a homogeneous area). The attributes of a surface can be expressed as continuous single-valued functions, or as step functions of independent continuous variables. These are commonly treated as either contour maps or sets of labelled regional boundaries. In both cases, irregular polygon boundaries provide an optimal data representation. Although the data contents of these two types of irregular polygon boundary are different, and the subsequent manipulation processes consequently differ, the same basic problems of graphic to non-graphic conversion and storage present themselves.

The range of non-graphic codes that can be employed to describe irregular polygons was generally discussed in Chapter 4 (pp. 87-90). Perhaps the most common form of regional coding used is that of topologically unrelated names referring to an external metric index (typically a map); the widely used names of countries, provinces, states, counties, parishes, precincts, traffic zones, administrative districts, and census subdivisions all fall into the category of named irregular polygons. The names have value for allocation and grouping purposes, are the basis of a vast amount of statistical literature, and are not in themselves metrically defined.

The relative positions of irregular polygons can, of course, be defined by use of adjacency codes or a connectivity matrix, but if any high degree of accuracy is required in the geographical location of the irregular polygon boundary, metric codes based on a geometrical coordinate system must be employed for non-graphic storage.

The complex shape of irregular polygons makes the task of spatial definition greater than that for lower orders of graphic representation (regular grids, points, or line segments). An additional impetus to precise line definition of polygons is that the location of the line along its length represents a boundary condition, whereas lines in networks representing flow connectivities, for example, can fulfil their function with less linear precision.

The volume of spatial data needed to define irregular polygon boundaries has considerable impact on the effort needed to convert the data from graphic to non-graphic format (which was the main constraint in current system design and use, as recognized in Chapter 4, p. 126). It also affects the storage and manipulation of the data, and the choice of file structures.

The relationship between the volume of storage data and the choice and resolution (fidelity) of the various storage formats is illustrated in the following series of diagrams (pp. 133-135).

Fig. 5.1 is a sample map (graphic image only) made up of an arbitrary set of lines within a 10-sq.in. rectangle. With different sets of descriptor data related to such a map, the lines could conceivably represent a network (such as a highway network) or a set of 15 bounded regions (such as a present land use map). There are approximately 33 in. of line length in the total set of interior and exterior boundaries. In comparison with real-world maps of similar subjects, the sample map might subjectively be described to be of "medium" data density.

Fig. 5.2 shows the result of regarding the sample map as a set of irregular polygons and using the point data storage format to record their centroid coordinates. Approximately 15 data points would need to be coded if this approach were used. Fig. 5.2.1 shows the size of a square grid that would result if 15 arbitrary regular area cells were used to describe the same area.

In Fig. 5.3, the sample map is taken to be a network where the flow between intersections is of concern and where it can be adequately represented by straight-line segments between intersections. (An intersection occurs where three or more line segments meet; map corner points are a special case.) Such a line segment map could be satisfactorily encoded with 32 data



Fig. 5.1. Sample map of 15 irregular polygons.



Fig. 5.2. Centroid coordinates (15 data points).



Fig. 5.2.1. Fifteen arbitrary regular-area cells.



Fig. 5.3. Network representation (32 data points).

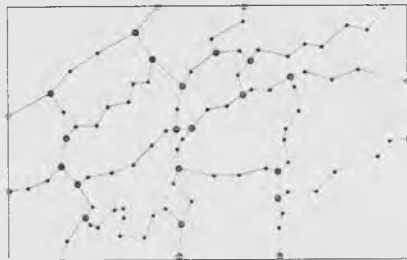


Fig. 5.4. Coarse polygon representation  
(98 data points).

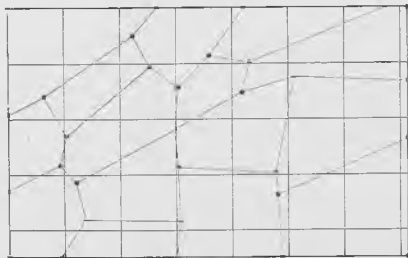


Fig. 5.3.1. Thirty-two arbitrary regular-area  
cells.



Fig. 5.4.1. Ninety-eight arbitrary regular-area  
cells.

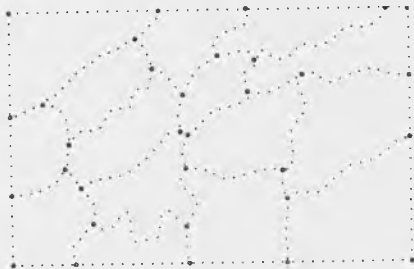


Fig. 5.5. Fine polygon representation: 10 points per in. (330 data points).

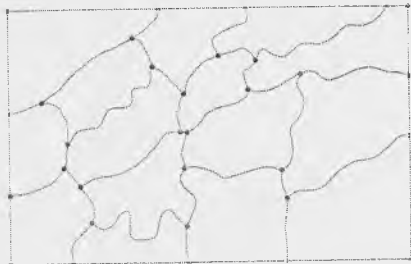


Fig. 5.6. Fine polygon representation: 50 points per in. (1,650 data points).

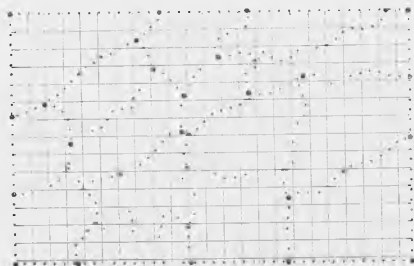


Fig. 5.5.1. Three hundred and thirty arbitrary regular-area cells.

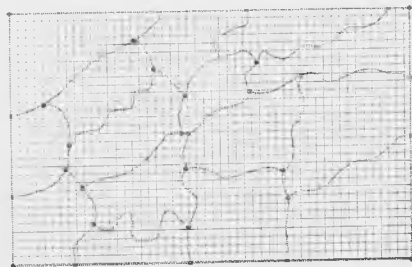


Fig. 5.6.1. One thousand, six hundred and fifty arbitrary regular-area cells.

points. Fig. 5.3.1 shows the size of square grid that would result if 32 arbitrary regular-area cells were used to describe the same area.

In Fig. 5.4, the sample map is regarded as a set of irregular polygons whose boundaries must be defined. Accordingly, the boundaries are represented, in coarse approximation, by straight-line segments between nodes (coordinate points) that have been manually selected between inter-sections. This is probably the lowest level of fidelity (and hence data volume) that would be acceptable to a user interested in the location of the boundaries. Ninety-eight data points are used to encode the boundaries in this manner. Fig. 5.4.1 shows the size of grid square that would result if 98 arbitrary regular-area cells were used to describe the same area.

Figs. 5.5 and 5.6 are topologically the same as Fig. 5.4; the difference lies only in the spacing of the coordinate points used to describe the configuration of the boundaries. In practice, the storage formats illustrated in Figs. 5.5 and 5.6 result from the use of various types of semi-automatic digitization techniques which produce a stream of coordinate points by following or scanning the lines on the source map. The spacing of the points is subjective and can be varied within one map. However, 10 points per in. is usually considered to be the widest spacing at which the configuration of irregular lines can be finely approximated. A spacing of 50 points per in. is common; it allows polygons to be defined with minimal loss of areal accuracy, but with lines that do not appear "smooth" to normal visual inspection. A spacing of 100 points per in. produces points that are within the width of the original line on the map, and this must be considered to be the maximum needed for accurate representation of the lines. Such a spacing might be considered necessary for cartographic purposes, where the locational accuracy of closely spaced lines depicting contours (for example) must be preserved. Given that there are 33 in. of line on the sample map, it would be necessary to encode 330, 1,650, and 3,300 points respectively for the spacings mentioned above. Fig. 5.5 illustrates a fine polygon representation using 10 points per in. (330 total points), and Fig. 5.5.1 shows the size of grid square that would result if 330 arbitrary regular-area cells were used to describe the same area. Fig. 5.6 illustrates a fine polygon representation

using 50 points per in. (1,050 points in total), and Fig. 5.6.1 shows the size of 1,650 grid squares to describe the same area.

Each of the approaches to data storage described above has advantages and disadvantages for data representation and subsequent data manipulation. It is important to note that a strong relationship exists between the storage format chosen and the volume of data generated. Metric definition of irregular polygon boundaries incurs the expense of storing significantly higher volumes of data than storage formats that do not attempt to represent the polygon boundary accurately. This is intuitively self-evident, but the degree of increase in data volume between formats is instructive. At a "coarse" level of irregular polygon definition (Fig. 5.4), the volume of data required is 3 to 6 times greater than lower-order formats. At a "fine" level of irregular polygon definition (Figs. 5.5 and 5.6), the volume of data required is 10 to 200 times greater.

The following series of diagrams (pp. 138-39) provides a general comparison between storage formats using irregular polygons and those using arbitrary regular areas. The volume of data and the comparative resolution of the two storage formats are of concern, where boundary definition attempts to be explicit.

The first figure (Fig. 5.7) is simply the sample map from the previous series, with a single irregular polygon indicated. The next five illustrations (Figs. 5.7.1-5.7.5) show the boundary of the same polygon as it would be stored if assigned to a regular arbitrary grid incurring progressively larger volumes of data storage. (The value of the irregular polygon is assigned to a regular grid unit if more than 50% of the regular grid falls within the irregular polygon.)

In each case, the arbitrary grid size chosen uses the same volume of storage as the alternative storage format on which it is superimposed. It can be seen that the storage of irregular polygons in arbitrary grid format either gives an unreal form to the real-world data element or imposes a higher burden of data volume to obtain comparable fidelity. It must be emphasized that this particular series of examples illustrates the effect



Fig. 5. 7.

Sample map with single  
irregular polygon indicated.



Fig. 5. 7. 1.

Polygon boundary assigned  
to regular grid using same  
data volume (15 data points)  
as centroid coordinates.

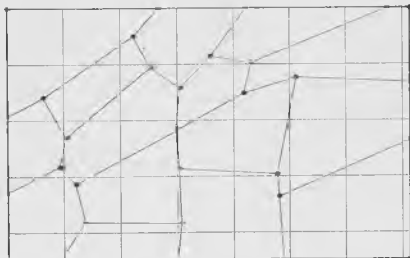


Fig. 5. 7. 2.

Polygon boundary assigned  
to regular grid using same  
data volume (32 data points)  
as line network.



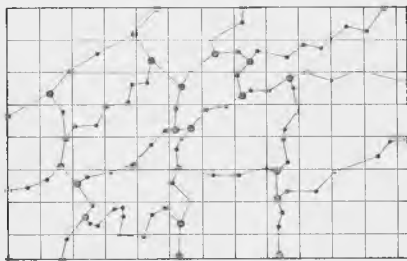


Fig. 5.7.3.

Polygon boundary assigned to regular grid using same data volume (98 data points) as coarse polygon.

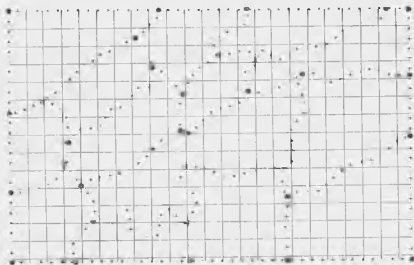


Fig. 5.7.4.

Polygon boundary assigned to regular grid using same data volume (330 data points) as fine polygons at 10 points per in.

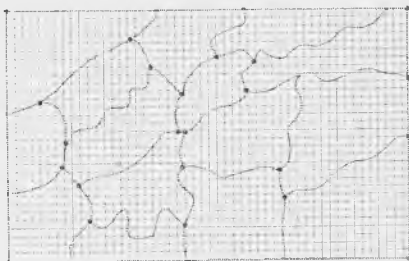


Fig. 5.7.5.

Polygon boundary assigned to regular grid using same data volume (1,650 data points) as fine polygons at 50 points per in.

only on one specific polygon. The error inherent in the transformation between storage formats may be less severe if the shape of the irregular polygon is closer to that of the arbitrary area used, and will probably be more severe if the irregular polygon has a more complex configuration. No measures of the degree of areal correlation can be validly obtained from this single example. The polygon used for comparison, however, is not a complex irregular polygon, and the general trend illustrated is probably valid. Given real-world data elements in the form of irregular polygons, a high volume of data is needed to define their boundaries metrically, but other types of storage format provide representation at a lower resolution even with equivalent volumes of metric data.

In consequence, only a relatively narrow range of non-graphic codes can be used to define the location of the boundary of an irregular polygon metrically without loss of resolution or prohibitively expensive (grid square) volumes of data. Within the range, three approaches can be recognized.

The first approach, and the most common, is the one used in the examples above, which can be referred to as "chain encoding". Coordinate points at chosen spacings are recorded either manually or by machine to define the boundary, with straight-line segments assumed to join the points.<sup>1</sup> A variation on the same theme is to define boundary points (usually intersections) and describe the point locus between intersections by a series of direction and distance codes, without actually recording the coordinate values of the interstitial points. The approach using coordinate point records is the one employed in the majority of irregular polygon systems and will be described more fully in the system descriptions given below.

The second approach, and perhaps the most mathematically elegant, is to approximate the boundary line piece by piece with analytically simple curves, as is done in SKETCHPAD and its successors.<sup>2</sup> This approach can greatly reduce the volume of data needed to define a polygon boundary metrically, at the expense, however, of some subsequent complexity in data manipulation. The approach is mentioned for the sake of completeness, but has not been adopted by any irregular polygon system now under development.

1. Freeman 1961, Freeman and Gardner 1964. 2. Sutherland 1963.

The third approach is the representation of planar regions by their "skeletons". If the skeletons, as defined by a parcel of grid squares, of adjacent plane regions are metrically described in sufficient detail, the boundary between them is thereby metrically defined.

Several types of skeleton representation can be envisaged; the one used by Pfaltz and Rosenfeld<sup>1</sup> assumes a rectangular matrix of points overlaid on the area containing the region. Within the matrix, "neighbourhoods" can be defined by a point and a radius.

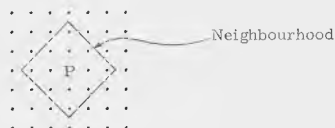


Fig. 5.8. Maximal neighbourhood of a point in a matrix.

In Fig. 5.8 above, the neighbourhood is defined by the location of point P with a radius of 2. A neighbourhood is termed a "maximal neighbourhood" if it is not properly contained in any other such neighbourhood. An irregular polygon can be regarded as a union of maximal neighbourhoods, and thus can be specified by centres and radii of those neighbourhoods.

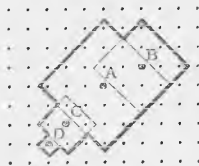


Fig. 5.9. Representation of an irregular polygon by maximal neighbourhoods.

The irregular polygon in Fig. 5.9 is defined by the maximal neighbourhoods of points A, B, C, D with radii 3, 2, 1, and 0.

Pfaltz and Rosenfeld compared the volume of data generated by this approach with that of chain-encoding irregular polygons, and came to the conclusion

1. Pfaltz and Rosenfeld 1967.

that for the same resolution of boundary definition approximately the same number of data points must be encoded. Specifically, for a series of irregular polygons (country borders), the results given in Table 5.1 were obtained.

Irregular polygon	A	B	C	D	E	F	G
Boundary points	338	178	231	286	440	197	262
Skeleton points	371	135	407	272	402	183	244

Table 5.1. Number of data points required to define a series of irregular polygons by chain-encoding boundaries and by skeleton method.

It is possible to generate a chain-encoded representation of a boundary, given the skeleton points, and it is more efficient to carry out some types of set-theory operations using such encodings. In particular, the method provides a very efficient form of storage if it is necessary to determine whether points lie within or outside regions. The approach also seems amenable to hierarchical coding of descriptor data, and would provide economies in some forms of generalization and retrieval of descriptor data. Unfortunately, the approach has not been adopted by any agency for an irregular polygon data handling system under development at this time.

The definition of irregular polygon data systems used at the beginning of this chapter was that the irregular polygons themselves must be defined in the primary file structure of the systems concerned. It can be seen from this discussion of metric definitions that polygon boundaries are generally defined with reference to some structure of coordinate points. In addition to carrying the required volume of data, if polygon forms are to be identified the file structure must obviously include coding of the relationships between the elements of the polygon form; between points that form line segments; and between line segments that form polygons. This "chain" must be explicitly or implicitly identified. Similarly, if more than one polygon is concerned, the relationship between polygons must be coded to produce maps. This can be implicit in metric coding, but advantages may be gained from specifying the relationships between polygons more directly. In

addition, the descriptor data (which may be very complex) has to be related to appropriate polygon boundaries at an efficient level in the hierarchy of relationships between coded image elements. For example, if many coordinate points are used to define a polygon boundary, it may be prohibitively expensive in terms of data volume and processing cost to attach the description of the polygon content to each point; whereas in a point data system this is a normal procedure. Clearly, irregular polygon data systems demand more sophisticated data structures than lower-order systems. Examples of the data structures employed in systems under development are given below.

Systems using the irregular polygon approach commonly use a tracing type of digitizer, an automatic line-following device, or some form of electro-optical scanner to produce the coded metric representation of the graphic source data. A large-scale computer is most often used to store and manipulate the data, but recent developments in mini-computers have allowed them to be used for creating and accessing large off-line data stores. Data manipulation capabilities affected by the format include the ability to measure areas and perimeter lengths, and to group data in sets of polygons either known to the system or arbitrarily generated by the system. Perhaps the most useful capability is that of overlaying one set of irregular polygons on another, to determine the characteristics of composite surfaces.

Polygon data systems have been adopted for handling surface-descriptive data sets from relatively large areas; the systems have commonly been developed to handle regional data. Data retrieved from such systems are typically displayed on a line plotter or in tabular form on a line printer, although the use of cathode ray tubes (both storage tube and refreshed display) is becoming more prevalent.

#### TYPES OF IRREGULAR POLYGON SYSTEMS

Irregular polygon data systems commonly use chain-encoding to define the polygon boundaries. For convenience of description, two types of chain-encoding can be recognized. The difference between the two types is related solely to the practice of encoding and utility of the subsequent records, as the approaches are topologically identical. The first relies on the

approximation of polygon boundaries by straight line segments between nodes visually chosen by an observer. An example of this approach is shown in Fig. 5.4 above; it results in records that will be termed "coarse" polygons. The second relies on the definition of the polygon boundaries by a stream of points (between which straight line segments are assumed), usually generated by a semi-automatic or automatic device. This approach (Figs. 5.5 and 5.6 above) results in records that will be termed "fine" polygons. This is an entirely arbitrary distinction and one of only temporary merit, as new encoding methods are being developed by which irregular polygons will be defined by metric codes at various points on the continuum between "coarse" and "fine". The categories do, however, reflect two main approaches to the design of irregular polygon data systems which have been attempted in the past decade and are employed by systems now under development. Several systems for handling irregular polygons in both of the above categories exist to store and manipulate contour data; however, the examples described below will be drawn from systems that use irregular polygons to define regions, as the problems are of comparable complexity and the regional systems relate more directly to the author's work with the Canada Geographic Information System (described in the Appendix).

#### Coarse Polygon Systems: the Example of PIOS

Several systems have been designed with coarse polygons as a primary data storage format. Typical examples are MAP/MODEL<sup>1</sup> (Map Model System, Columbia Region Association of Governments, Portland, Oregon), PIOS<sup>2</sup> (Polygon Information Overlay System, Comprehensive Planning Organization, San Diego County, California), and NRIS<sup>3</sup> (Natural Resources Information System, Boeing Computer Services, Inc.). All the systems exhibit similar basic characteristics and closely follow the concepts that originated in the MAP/MODEL system between 1964 and 1968. Despite the length of time over which the systems have been developed, recent communication with those responsible for their development confirms that the approach is still experimental. None of the systems is in routine operation in an agency that is using, rather than developing, the approach. The PIOS system is briefly described below.

1. Arms 1967, 1968, 1969, 1970. 2. Dangermond 1971, Comprehensive Planning Organization 1971. 3. Boeing 1972a, 1972b.

Like MAP/MODEL and NRIS, PIOS is based on the concept that the graphic content of a map is made up of irregular straight-sided polygons that define homogeneous areas. Points, and line segments that do not bound polygons, are considered to be special cases of polygons.

The purpose of implementing PIOS in San Diego County, California, is primarily to summarize soil data by traffic zone for use in an urban development model. The secondary objective is to provide a capability to aggregate the soil data into other areal units economically for subsequent applications.

Two sets of maps have been specially prepared for the project. All maps are drawn on stable base material (mylar) at a scale of 1:24,000 with an approximate sheet size of 40 in. x 45 in. The state plane coordinates of at least three reference points (edge "tick" marks) are given on each sheet, and each sheet is uniquely numbered within its set. The first set of approximately 15 sheets (some partial sheets) shows soil data boundaries (irregular polygons). Each sheet is overlaid on a separate identical copy (blueprint) on which are written the soil-type symbols. The total soil data map set contains approximately 30,000 polygons. The number of polygons on each sheet varies widely, but subjectively can be considered to be of medium to low density (a range of 500 to 5,000 polygons per sheet). The second set of maps are overlays for the first set and show traffic zone boundaries (irregular polygons); the total of such polygons in the map set is approximately 250. Again, the number of polygons on each sheet varies widely, but all can be considered to contain low-density data. The traffic zone number is written directly on the overlay. The first step in the overall project is thus the establishment of a totally graphic storage file.

The boundary data are metrically coded and transferred from graphic to non-graphic storage by using a digitizer to record a series of discrete points that define each polygon. The area of each polygon is considered to be completely enclosed by a border composed of straight line segments joining the points. (Curved lines are approximated by such straight line segments.) Each individual polygon is totally enclosed by a series of points, so all lines on the map are recorded twice. The positions of the points are visually selected by the operator; the number required to define a polygon varies with

the size and configuration of the polygon. The finest resolution used is approximately 50 points per linear inch of polygon boundary, but the average is substantially less than that. A total of 6,000,000 points have been used to define 30,000 polygons, an average of 200 points per individual polygon record. The record size, however, varies widely. Each polygon is also assigned the coordinates of a contained centroid visually estimated by the operator.

In addition, several significant items of data are added to the content of each polygon record. Each polygon is assigned a unique number (a combination of the map number and the polygon number within the map). Closest polygons are numbered consecutively to achieve some efficiencies in subsequent grouping and processing procedures.<sup>1</sup> The descriptive data relating to the polygon are entered as part of the polygon record. (Soil data alphanumerics are transformed to a numeric code so that they can be entered through the digitizer keyboard.) Each polygon is assigned a code by the operator to distinguish normal polygons, incomplete polygons (those at map borders), and polygons wholly contained within other polygons (the island case). A count of the total number of points required to describe the polygon is automatically assigned to each polygon record by the digitizer.

For each sheet, the coordinate axis of the digitizer table is manually aligned to the axis of the Plane Coordinate System reference points on the map to eliminate the need for subsequent calculation of coordinate rotation. The table coordinates of the reference points are recorded for each map, the upper left reference point being assigned a table coordinate value of 0.0.

The effort expended in digitizing alone, excluding error checking and redigitizing, took 218 man-days and cost \$10,240.00, excluding overhead. These figures reflect the level of density and complexity of configuration of the source data. Hardware problems also contributed to the amount of effort involved (though these are not unusual in this type of development). Dangermond<sup>2</sup> estimates that overall digitizing costs could be lowered by 30% in a production operation.

An inspection of the spatial connectivities inherent in the data records

1. This is a loose form of topological connectivity coding, but is not rigorous enough for subsequent use as a connectivity indicator. 2. Dangermond 1971.



produced by the digitizing process makes it possible to understand the subsequent processing steps used to produce an internally consistent non-graphic file. As a result of the digitizing process, points are specified by metric codes and their spatial relationship to one another is defined by their sequence within the line segment records. Similarly, the relationship of the line segments (the straight line connection points) to one another is established by their sequence in a polygon record (and only one polygon record). Spatial connectivity between polygons is explicit only in the metric codes of the boundary points used to define them. The centroid coordinates or the unique number assigned to each polygon would seem to be amenable to topological connectivity coding, but none has been attempted. Map-to-map connectivity is loosely defined by the map numbering system, but is explicit only in the coordinate values assigned to the edge reference points. The specification of spatial connectivity thus lies within the metric coding of the data set. The spatial aspects of subsequent file creation are initially concerned with achieving an internally consistent metric structure. Such a system should allow editing and correction of errors in the coordinate records. Assuming that a complete map is not digitized at one time, then there must be a way to include polygons in one set to form a complete map or set of maps. Assuming that more than one map is digitized, there must be a process to link the metric codes of incomplete polygons to form whole polygons (edge matching). If subsequent processing is to be efficient, there must be a process to convert the coordinate description of the file to a uniform base. Subsequent data manipulations such as data retrieval, measurement, and comparison will probably need to be based on the metric properties of the file. This indeed occurs in PIOS.

After the digitizing process is completed, some initial steps incorporated in the programs STORE AND EDIT, POLYGON MERGE, and MAP MERGE are gone through to facilitate subsequent processing. With these programs, each polygon record is edited for such items as closure, format, and correct contents (none missing or exceeding given limits). In the process, the extremity coordinates of the polygon are determined and added to the polygon record. The polygons are plotted using an automatic drafting

machine under control of the computer and the resulting graphic representation of the digital record is visually inspected to determine inaccuracies, particularly non-logical errors such as those resulting from digitizer drift, incorrect set-up, or others. (These problems will be discussed more fully in a later part of this chapter.) Inaccurate polygons are redigitized and the new records are exchanged for the old on the basis of polygon numbers. Finally, the polygon coordinates are changed from those of the digitizer table to state plane coordinates, by employing the state plane coordinate values of the reference points on the border of each map. The file that results from these operations is termed the Base Map Master File.

The next series of operations is designed to remove the base map borders, construct complete records for polygons that contain others, compute areas and centroids, convert polygon vertices to centroid relative coordinates, and create the PIOS Master File with amplified descriptive data. These functions are performed by the interaction of three programs (LIST, CONNECT, and INSERT). The LIST program is used to produce a list of the numbers of the border (incomplete) and nested (island case) polygons from the sheets being joined. The listed numbers of the appropriate polygons are matched by visual inspection of the graphic source material. Thereafter, a temporary file is generated which contains the results of this visual process in the form of an indexed sequential set of polygons. One pass of the program CONNECT uses a search radius technique to match coordinate strings in adjacent border polygons from adjacent maps. The same approach is used to insert contained (nested) polygons into the file by a manual match followed by machine merge, though in this case the merge is achieved using the INSERT program. The INSERT program also calculates the area of each polygon (using Simpson's rule), and calculates a "centroid" (using a straightforward centre-of-gravity solution; the resulting point is expressed in state plane coordinates, and may be within or outside the polygon). The same program converts the polygon vertex coordinates to centroid relative coordinates. This latter step facilitates certain subsequent manipulations, including a simplified version of projection change, using such programs as those available in the AUTOMAP system but performed on the calculated centroids rather than on all vertex coordinates. In addition, the INSERT program

writes literal record subtitles into the descriptive record, such as SOILS, TRAFFIC, CSPC (California State Plane Coordinates), as appropriate in each case. The resulting file (PIOS Master File) is thus a sequential set of polygon records, each record having an indexed sequential structure, with a relatively sophisticated set of inserted and calculated data contents.

The following example is an indication of the level of effort required for an average map to be processed through the above steps from digitization to PIOS Master File. The map used as an example contains 3,100 soils polygons. The times given are for single runs of each of the programs. The runs were made on an IBM 360/50, 256K, with a Calcomp 566 30-in. plotter.

<u>Computer procedures</u>	<u>Run times</u>
1) Store and editing (several runs to get total map)	17.3 min cpu
2) Merge program (several runs to get total map)	19.7 min cpu
3) Plotter runs	8.1 min cpu 4.2 h plotting time
4) Editing run	8.7 min cpu
5) Convert run	37.7 min cpu

Table 5.2. PIOS: computer time required to process an average map from digitization to Master File format.

Several kinds of manipulation can be performed on the PIOS Master File. The LIST program can be used to list and summarize polygons by type, which can be specified by any of the characteristics in the polygon record (see above). This represents a valuable capability for selective retrieval. The PIOS Master File can be updated using the INPUT program. By manually specifying the changes on new punch cards, it is possible to add, replace, or delete polygons, and to change descriptive information. However, this may require laborious manual preparation of punch cards unless the changes involve small amounts of data. The MASTER FILE PLOT program will cause the automatic plotting table to draw maps of the polygon file or portions of the file (see "window" capability below), at any scale, suitably gridded and titled. Descriptive annotations are placed at the estimated centroid contained within each polygon (rather than at the calculated centroid). Weighted shading can be assigned to specified areas. An alternative form

of graphic display is to allocate descriptive data of the polygons to their centroids and use the resulting point data as an input to systems of the SYMAP type; in other words, a grid cell data bank of any cell size can be created. Similarly, such point data could be used in any point data handling system, could be aggregated by point-in-polygon techniques, and could be used in comparative statistical or single variable analysis. These modal change procedures do, of course, incorporate all the limitations of the resulting data formats discussed in Chapter 4, pp. 95-97, although they have some value for planning studies that involve small-scale grid techniques.

More interesting are the "windowing" and "overlay" capabilities. The windowing technique allows rectangular areas to be sorted out of the master file for subsequent computation and analysis or graphic display. The rectangular area required is specified by coordinates and the program creates a work file that contains those polygons whose extremities fall within the rectangle, that is, within the specified "window". This limits the size of file to be accessed for subsequent processing, particularly for overlay purposes.

Overlaying is the process of superimposing two or more data sets such that the final data set produced contains the information from both or all of the original data sets. As the number of overlays and map complexity grow, so do the cost and doubts about reproducibility. This is obviously a very expensive and time-consuming process for complex situations, although it may be acceptable for simple ones. The overlay process employed in PIOS is based on several assumptions. The data sets to be overlaid are, of course, sets of defined, straight-sided polygons. It is assumed that the average size of one of the polygons of one set (the traffic zone) is substantially larger than the average size of polygons in the second set (the soil maps), and that the latter are to be aggregated within the former. It is also assumed that the data sets are of limited size (this is ensured by the ability to "window" the required areas), and that only two sets of data are being overlaid at one time.

Various approaches to the task of overlay are available.<sup>1</sup> The approach chosen by the designers of PIOS<sup>2</sup> reflects the necessary balance between

1. Arms 1968. See review of overlay techniques, pp. 795-805 in Tomlinson, R.F. 1972. 2. Dangermond 1971.

efficiency of processing and the core storage requirements needed to run the program, given the overlay assumptions mentioned above. The algorithm used employs an efficient polygon subdivision technique which partitions the larger polygons for subsequent point-in-polygon calculations to determine which vertices of the subordinate polygons are contained within them, and hence which parts of the subordinate polygons are contained. Each dominant polygon (traffic zone) is divided into 16 strips of equal width, parallel to the X axis. Each strip is thus a newly defined sub-polygon, which is stored in an array. Each point in the subordinate polygon is read into the array and is tested to determine whether it is in or out of this strip of the major polygon. As the point-in-polygon technique operates parallel to the strip's longitudinal direction, the test need only be applied within the strip and the rest of the dominant polygon need not be considered. After the special case of locating the first point has been considered, successive points define line segments that are contained within, intersect, or are outside the dominant polygon. Those contained are recorded in a memory table. Those intersecting allow the intersection point to be calculated, the truncated line segment to be stored in the memory table, and a clockwise change in direction to be made to proceed to the next point. In this manner the new (overlaid) polygons are developed and their areas are subsequently computed. The resulting polygon sizes are summarized by type within the dominant polygons.

This is an extremely straightforward approach to polygon overlay, and the preliminary assumptions made and described above eliminate the majority of problems inherent in more complex polygon overlay tasks (such as image blackening, small area creation, excessive data volume creation, coinciding boundaries, coastline weave, and immersed boundaries). The few problems that arise are visually examined in a graphic plot of the file resulting from overlay and are corrected manually. In consequence, the computing times for the overlay processes are short. The simplest case, where two polygons are overlaid one on the other and have two intersection points, has an average computation time of 0.8 sec of cpu time. A typical window containing 400 soil zone polygons overlaid by one traffic zone can be analysed

in less than 1 min of cpu time, or an average of 6.6 polygon comparisons per sec. (A considerable percentage of the 400 polygons involved in such an example will not be included in the dominant polygon and hence will not be processed.) The overall effort to overlay all of both data sets in the PIOS project cost \$4,200.00 for computer time and \$130.00 for labour, and was completed in 2 days.

#### Fine Polygon Systems: the Example of CGIS

Probably the most comprehensive system designed with fine polygons as a primary storage format is the Canada Geographic Information System (CGIS).<sup>1</sup> Although the system is currently in operation at the Department of the Environment, Government of Canada, Ottawa, the system facilities are under continuing development. A full technical description of CGIS is given in the Appendix.

The approach used in CGIS is, first, to transform the graphic data into a structured, non-graphic format amenable to computer processing, and secondly, to provide a set of procedures that can efficiently operate on the stored data. These two steps are conceptually distinct. The creation of the data bank involves the processing of source data by a uniform set of operations. Only the data, or at most a few parameters, change. This set series of operations, or "phases" as they are referred to in the CGIS, remains constant and is executed under the resident operating system of the computer. The subsequent manipulation and retrieval of the data, on the other hand, are made up of sets of processes that can vary for each user and each use. The sequence and type of operations may change with each request, for example, overlay, dissolve, or merge. Each request has to be provided with its own "job stream", which can be complex to set up, time consuming, and prone to error unless automated. Accordingly, a "monitor" system is a basic need of the retrieval process. It facilitates and coordinates the operations needed to manipulate the data bank, and also has a special compiler that allows a number of commands for graphic data handling to be expressed in an English-like language. The output from requests can be in tabular or plotted line form.

1. Tomlinson, R. F. 1967, 1968.

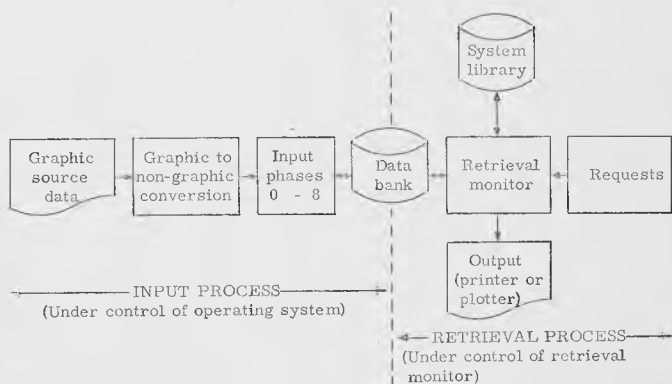


Fig. 5. 10. CGIS: overall system concept.

The first function of the system is to convert the graphic source data to a simple non-graphic format. The operations involved in this process are illustrated in the diagram (Fig. 5. 11) below, but it may be noted that the graphic source data are provided as hand-drawn annotations on translucent, stable-base reproductions of topographic maps. The volume of source data is high. Over 15,000 20-in. x 30-in. sheets have already been received for system processing. The source data are prepared for coding by tracing (scribing) the image (boundary line) data on separate sheets, by sequentially numbering<sup>1</sup> each graphic data element on each sheet, and by creating a sequentially numbered list of the alphanumeric descriptors that apply to each graphic data element on each sheet.

Thus prepared, the image data are scanned with a drum scanner<sup>2</sup> (see Fig. 5. 11) to create a magnetic tape of the image. The sequentially numbered alphanumeric descriptors are keyed<sup>3</sup> directly on magnetic tape. To link the image and descriptor data sets, a digitizer produces a third magnetic tape that contains the coordinates of a point for each sequentially numbered data element (usually a visually estimated, contained centroid for polygon

1. This process is duplicated to allow data to be verified during subsequent editing procedures. 2. Thompson 1967. 3. This process is duplicated to allow data to be verified during subsequent editing procedures.

data). The number is common to the point and to the alphanumeric descriptor of the data element. The coordinates of the point can be metrically related to the coordinates of the image record. A record is also made of certain data, such as map scale, latitude and longitude of map corner points, and type of source data, to be used as control information in subsequent processing.

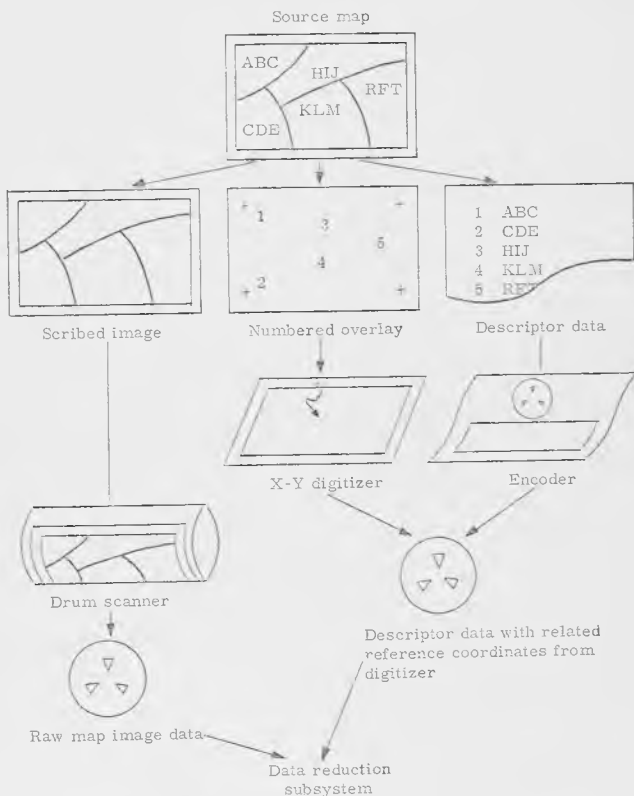


Fig. 5.11. Operations involved in CGIS process of graphic to non-graphic conversion.



Once the data are in non-graphic form, the following system functions are carried out by a computer.

The non-graphic data, both image and descriptive, are edited for validity.

Redundant data are eliminated and topological codes are automatically added to the metrically coded images. Line segments are identified as line segments, are ordered and listed, and are assigned "directions", that is, arbitrary but constant "right" and "left" sides. Single-bounded polygons are identified as polygons. The connectivity between polygons and line segments is specified in a code automatically assigned to each line segment record.

The metric codes of the image records are converted from the coordinate systems of the scanner and digitizer to a uniform base coordinate system, the Geographical Coordinate System (GCS). When applicable, the input scale of the source data is converted to a storage scale, and if this involves scale reduction, single-point smoothing of the lines is carried out. The effects of linear distortions in the source material and axial rotation on the absolute position of data elements are eliminated during these transformations. The area of polygons is measured and the centroid coordinates (centres of mass) of the polygons are calculated and added to the polygon records. The size of the image data file is reduced by converting the metric codes for line segments to a compact format. The file containing the image data is automatically subdivided and labelled in conveniently sized blocks (frames) for ease of subsequent retrieval and processing; the system thus imparts a structured areal division to the image file. Polygon equivalences are resolved and the descriptor data are matched to the image data.

The system automatically detects topological errors and lists the errors found in the files. A manual subsystem allows errors to be corrected by using manually prepared correct data elements to update the file.

After error correction of the files, edge matching is achieved between the new sheets of data and the data already stored in the data bank. The new data are then added to the data bank (data bank creation).

The conceptual format of the data bank resulting from these processes is shown in Figs. 5.12 and 5.13.



of the image data needed to describe the boundary of one map element described in the DDS.

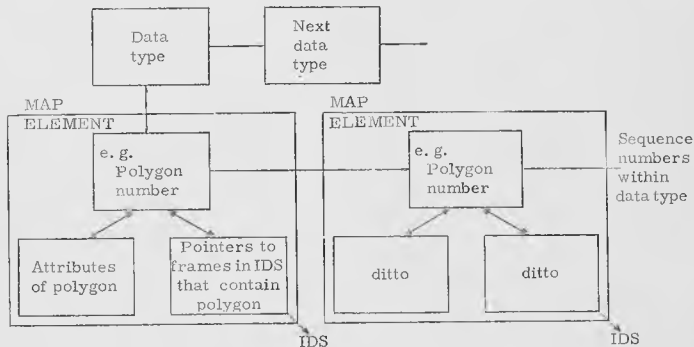


Fig. 5.13. Descriptor Data Set - Level 1.

Various additional levels of structure can be imparted to the DDS by grouping and indexing the map elements; the subdivision of a file by region is a typical example. The resulting format is shown in Fig. 5.14.

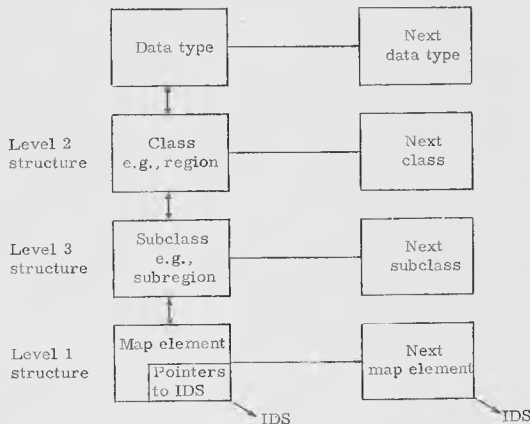


Fig. 5.14. Structured Descriptor Data Set.

As noted above, all functions of data bank manipulation and retrieval are carried out under the control of a retrieval monitor program. One of the basic system capabilities is thus the ability to combine manipulation or retrieval commands, or both, automatically, with the descriptive information in the appropriate file and with computer programs resident in the system library. It generates a complete request that the computer can use to enter the data bank and produce the desired answer in tabular or plotted form. The set of code words used to specify the manipulation and retrieval functions make up the "command language".<sup>1</sup> The manipulation and retrieval functions currently in use are described below. In each case, the command language word or words that initiate the function are given in parentheses.

Within any one data type (coverage in the data bank), it is possible to retrieve any desired subset of data. Specifically, one can select (extract) all data with desired IDS or DDS characteristics and allocate a name to the resulting subset (SELECT, INCLUDE, NEWNAME); select all data except those with undesired characteristics and name the resulting subset (SELECT, EXCLUDE, NEWNAME); or specify a circle or polygon within which to address further commands (CIRCLE, POLYGON).

The structure of a file in the data bank can be changed, by sorting the DDS into a desired sequence of any characteristic (DDSORT); by specifying (labelling) parts of the file for further inclusion or exclusion in processing (COMBINE, MERGE); by equating existing faces in the file and effectively removing the boundary between them, so that for the purposes of further retrieval they will be regarded as one area (DISSOLVE).

The content of a file in the DDS can be changed by reclassifying existing map elements in the file. This allows the addition and deletion of data (DISSOLVE XXXX FROM YYYY POINTER, CLASSIFY XXXX FROM YYYY POINTER). Reclassification of the IDS is carried out by using the manual process of error correction and reprocessing through input phases 0 to 8 (as detailed in the Appendix).

"Measurements" can be carried out on the file by gathering together the measurements already made in the input phases and converting them as desired to units of either a square mile or an acre (SQ MI, ACRE).

The content of a file can be selectively reported upon, printed out, or plotted. The whole file, either IDS or DDS, can be printed out (IDS DUMP, DDS LIT). Format control of such a printout or selected retrieval of part of the file, or both, can be specified (ASSESS). Any specified part of the IDS (with or without the DDS) can be plotted. The plot can incorporate a scale change within a range of  $\times 0.5$  to  $\times 4$ . Numeric labels can be plotted at the centroid of each face and the appropriate descriptors from the DDS can be listed separately by label number (PLOT).

From multiple data types (coverages), it is possible to extract (move) both IDS and DDS data from the data bank and place them in the system library in preparation for further combined processing (COMBINE). After this step has been taken, an overlay can be made of overlapping data sets (OVERLAY). The resulting single coverage is amenable to all the manipulation and retrieval functions mentioned above.

Several "housekeeping" functions can also be carried out. It is possible to modify the system parameters in the data bank and in the system library, create back-up files, and list the existing system library (MODMCCE). A mock DDS can be created for testing purposes or for generating a subset of data related to a circle or polygon (DDSGENR). New command words can be specified and added to the command library (SYSMAT).

It can be seen from the above description of CGIS functions that the costs of handling spatial data in non-graphic form can no longer be simply regarded as those associated with converting the source data to non-graphic form plus those incurred in retrieving or deriving measurements and comparisons from the data. At least three discrete steps can be recognized that have a cost associated with them. The first remains the graphic to non-graphic conversion, although, as can be seen, this may incorporate several distinct processes, the contributive cost of which may vary with each data set. The second step is data reduction and data bank (file structure) creation. This step can be carried out within the computer but may incorporate several high-order operations of measurement and comparison, the results of which are stored in the file for subsequent retrieval. In the case of the CGIS, this step involves change of map projection, centroid calculation, area measurement,

linear measurement, automatic assignment of names to facilitate retrieval, topological editing and correction, and simple forms of line smoothing; it may also involve scale change, merge, and dissolve. The third step is the actual manipulation, retrieval, measurement, and comparison of data in the file. This may involve the retrieval and grouping of any of the measurements carried out during the creation of the data bank, or the generation of new measurements and comparisons, or both.

The task of directly comparing machine-aided techniques and equivalent manual processes can thus be extremely difficult and misleading. For comparison with one manual operation, the total costs involved in creation of the whole data bank represent a very large overhead burden from which it is impossible to separate the component parts that have contributed to the equivalent computer-aided operation. For the same manual operation, if retrieval costs only are examined the result may be equally misleading, as much of the necessary processing may have been accomplished in creating the non-graphic file and the retrieval costs may seem unreasonably low.

Let us, however, examine the costs for the various stages of data handling in the CGIS, to gain an initial overview of the effort incurred.

The most recent information on costs of using the CGIS is some test results made available in September, 1973. Table 5.3 below is a summary of the costs from which some conclusions may be drawn. First, let it be said that the test did not involve either a wide range of data types, or enough volume of data from which to derive reliable estimates for all types of data processing using the CGIS. Such a volume test is currently being designed, but will not be available in time for inclusion in this thesis. The test results thus refer only to the type and volume of data being processed during the test alone. (However, the data were chosen to be representative of types commonly available to the system.)

The source data for the test comprised four sheets of irregular polygon data with simple related descriptor data (a two-digit or three-digit descriptor code for each polygon). The sheets varied in size and data density. The sheet dimensions, polygon count, and total line lengths for each sheet are

given. The source data were provided by the United States Geological Survey and were based on U. S. topographic sheets. For the convenience of the CGIS, the sheets were redrawn to conform with Canadian National Topographic sheets, giving a total of 12 sheets. The resulting dimensions are tabulated below. The polygon count and the scale remain the same, the line length is adjusted to accommodate the additional length of sheet boundaries only, the data density is approximately halved, and the change in polygon density per square inch and line inches per square inch are tabulated. The figures relating to the source data are provided in the top four rows of the table, and those relating to the Canadian equivalent sheets are given in the bottom four rows; the rows are arranged in descending order of data density in both cases. The sheet containing the densest source data has 2.54 polygons per sq. in. and 4.02 line in. per sq. in., which might be subjectively described as medium-density data. The sheet with the least dense source data has 0.13 polygons per sq. in. and 0.68 line in. per sq. in., and this might be subjectively described as low-density data. In the Canadian format, the data have a range of 1.34 to 0.07 polygons per sq. in. and 2.13 to 0.40 line in. per sq. in. The costs given are those for processing the latter 12 sheets in the Canadian format. A moderate additional overhead burden is thus placed on the process, in that edges have to be matched between the sheets that make up the new data set, but the thematic data set being processed remains essentially the same.

The costs of graphic to non-graphic conversion include all processes between the source document manuscript and the production of a digital record suitable for computer processing. This involves the tasks, described above, of reformatting, image scribing, polygon numbering, descriptor data encoding, reference coordinate digitizing, and image scanning. The costs for this step are given in terms of total labour and total man-hours used, and the number of minutes taken to complete all these functions per square inch of map sheet are given in the adjacent column. It is interesting to note that the total number of minutes per square inch (with one exception) is less than 3, the level recognized in Chapter 3, p. 78, as being efficient for manual processing. This is to be expected, as the manual tasks involved are basically low-order operations such as tracing and numbering, and no data inherent in the file structure have to be measured or compared.

	TEST DATA CHARACTERISTICS							PROCESSING COST SUMMARY				TIME & COST ANALYSIS					
	Sheet dimensions (in.)	Perimeter (in.)	Sq.in.	No.of poly-gons	Poly-gons/sq.in.	Total line length (in.) incl. peri-meter	Line in. / sq.in.	Graphic to non-graphic conversion	Min. sec.	Data reduction and file formation	Total	Total cost/sq.in. (\$)	Total cost/polygon (\$)	Total cost/line in. (\$)	Data red. cost as manual equiv. time* (min./sq.in.)	Total cost as manual equiv. time* (min./sq.in.)	
Source data	18 x 26	88	468	1190	2.54	1833	4.02	Assumed as below	5.4	Assumed as below		1.76	0.68	0.43	15.7	21.1	
	38 x 44	164	1672	1252	0.75	3728	2.23		2.4			0.66	0.88	0.29	5.5	7.9	
	18 x 26	88	468	185	0.40	566	1.21		2.4			0.71	1.00	0.59	6.1	8.5	
	38 x 44	164	1672	207	0.13	1134	0.68		0.7			0.38	3.04	0.55	3.7	4.4	
<div></div>																	
As processed in Canadian format	Reformat (No. of sheets)	Total exterior perimeter sheet group (in.)	Approx. total sq.in.					Man-hr. Govt. of Canada labour costs		IBM 370/165 CPU min. gross system utilization costs							
	2	122	900	1190	1.34	1917	2.13	42.30 hr. \$182.55	2.4	24.092 min. \$624.92	\$ 806.47	0.90	0.68	0.42	8.4	10.8	
	4	217	3000	1252	0.42	3781	1.26	\$4.84 hr. \$204.85	1.3	21.491 min. \$519.99	\$1095.84	0.37	0.88	0.29	3.1	4.4	
	2	122	900	185	0.21	600	0.67	\$18.45 hr. \$83.25	1.2	4.624 min. \$254.08	\$ 334.28	0.37	1.80	0.53	3.2	4.4	
	4	217	3000	207	0.07	1187	0.40	\$20.25 hr. \$102.45	0.4	10.959 min. \$227.53	\$ 627.95	0.21	3.04	0.53	2.1	2.5	

\* Assuming labour costs of \$5.00 per hour.

Table 5.3. CGIS test: summary and time and cost analysis.



The costs of data reduction and file formation (data bank creation) are given in the next column. The cpu minutes give an indication of the relative amount of effort involved. They include all cpu utilization for phases 0 to 8, inclusive, plus MEC (error correction) procedures. Perhaps the costs of gross use of the IBM 370/165 system give a better indication of the level of effort. The costs are those paid to a commercial data centre for use of all facilities related to these stages of the processing, and not simply to cpu use.

The combined costs of all data processing from source data to data bank are given in the next column. As mentioned above, and described in detail in the Appendix, these costs include high-order techniques for handling graphic data such as area measurement, linear measurement, scale change, projection change, topological code assignment, and restructuring of the graphic file. The analysis of these costs in the next three columns shows what might be intuitively expected. First, the denser the data on the source material, the more effort (time and cost) per square inch is required to process them. Secondly, as fewer polygons per square inch are encountered, the cost per polygon rises (as both the size of the polygons, and the number of square inches they encompass, increase). Thirdly, the cost per line inch of data processed shows no marked trend. (The sample is not large enough to show a trend when differences are small.) Variation between examples in this sample are probably a function of the number of errors within any one data set that had to be corrected, or the edge matching problems encountered.

Perhaps the most interesting observations can be made from the time analysis. Obviously, in any machine-aided data processing system that incurs costs, the question arises as to the amount of manual effort that could have been brought to bear on the same tasks if equal amounts of money had been spent on men as opposed to machines. This is an extremely difficult question to approach, as so many variables are involved and so many assumptions must be made. In an attempt to provide some comparison, however, the total costs involved were converted to hours using an arbitrary figure of \$5.00 per hour for labour costs. (No attempt will be made to defend this figure. It is probably fractionally high in terms of straight salary, and substantially low if labour overhead is included.) The resulting man-hours are given as

minutes of labour per square inch of sheet processed. The range varies directly with the data density, as might be expected. With regard to the Canadian format (with its lower data densities), the range is 2.5 to 10.8 min per sq. in. (or 2.1 to 8.4 min per sq. in. if the low-order manual processes involved in the graphic to non-graphic conversion are extracted); with regard to the higher-density source data the range is 4.4 to 21.1 min per sq. in. (or 3.7 to 15.7 min per sq. in. if the low-order manual processes are extracted).

Direct comparison between the computer processing and manual-cost equivalent times for the complete range of data densities is given in Table 5.4 below.

Polygons per sq. in.	Manual equivalent (min. per sq. in.)	Data density	Line ins. per sq. in.	Manual cost equivalent min. per sq. in.
2.54	21.1	Medium	4.02	21.1
1.34	10.8		2.23	7.9
0.75	7.9		2.13	10.8
0.42	4.4		1.26	4.4
0.40	8.5		1.21	8.5
0.21	4.4		0.68	4.4
0.13	4.4		0.67	4.4
0.07	2.5	Low	0.40	2.5

Table 5.4. CGIS and manual equivalent times for handling data at various densities.

An interesting observation is that these time figures (a range of 2.5 to 21.1 min per sq. in.) represent the manual-cost equivalent minutes for completion of several high-order data manipulation processes. The range for single high-order data manipulation processes determined from the examination of manual tasks in Chapter 3, p. 78, was between 3 and 65 min per sq. in. In these terms, and from this very limited sample, it might be suggested that the machine-aided processes were somewhat more efficient than manual techniques, even when the step of graphic to non-graphic conversion is included in the costs.

Perhaps even more significant than the cost factor is the total elapsed time required to carry out equivalent tasks. Table 5.5 shows the time required to create the data bank, assuming actual time for manual procedures

(plus a nominal one-day allowance for machine process turnaround), compared with the number of man-days that could be purchased with the same funds.

Data density per sq. in.	Sq. in.	Elapsed time in CGIS days	Total CGIS cost expressed as man-days @ \$5 per h	Total real elapsed time for equivalent task
2.54	460	6.29	20.57	Based on manual estimates in ch. 4, these figures could be multiplied by a factor of 1 to 6 for real-time estimates
0.75	1672	6.72	27.52	
0.40	468	3.31	8.29	
0.13	1672	3.53	15.33	

Table 5.5. CGIS test: total elapsed time and cost required to create data bank.

The above analysis relates to the creation of the data bank and, although several high-order processes of data manipulation have been accomplished, the retrieval of the data has still to be carried out and paid for. This retrieval can, of course, include further measurement and comparisons performed on the non-graphic storage.

Before presenting these retrieval figures, it is pertinent to remember that a basic objective of this study was to compare the utility of non-graphic storage compared with the utility of graphic storage. A case could be made that the only costs relevant to such a comparison would be the ones presented below, and that the costs of creating the non-graphic storage given above should be regarded as overhead, just as the costs for creating the original graphic storage are generally considered to be overhead. The time has not yet come, however, when data gathering commonly proceeds to the non-graphic format without passing through the graphic phase; thus it is still reasonable to include the costs for the step from graphic image to data bank, particularly as several high-order processes of data manipulation are involved in creating the data bank. However, it should be emphasized that data bank creation is a one-time effort and represents an overhead cost that must necessarily be spread over all subsequent operations of data retrieval, much as survey and cartographic costs are spread over all subsequent processes of graphic data extraction (if they are considered at all!).

Several retrieval operations were performed on the data sets described above. Specifically:

- 1) Linear measurements of the perimeter of each polygon were made for each data set. These were expressed as inches of line and also converted to metres at map scale. The result was printed in a report, giving the code number of each polygon, its descriptive data, its size in acres, and the two lengths measured.
- 2) The area of each polygon was measured and expressed in acres, and the centroid of each polygon was determined and expressed in degrees of latitude and longitude. The results were printed as a report, giving the code number of each polygon, the latitude and longitude of its centroid, its area, and the percentage of the total area of the data set represented by that polygon.
- 3) A summary report was generated showing the sum of acres of polygons with the same descriptor, the number of polygons that made up the total, the average size of each polygon, and the percentage of the total data set of each descriptor.
- 4) An overlay of data sets of the same size was performed, that is, the medium-density data set of 468 sq. in. (1,190 polygons) was overlaid on the less dense data set of 468 sq. in. (185 polygons). Similarly, the 1,672-sq. in. (1,252-polygon) data set was overlaid on the 1,672-sq. in. (207-polygon) data set. In each case a report was generated showing the size in acres of the polygons formed by the overlay process, their linked descriptors, and the percentage of each new polygon contained within the less dense polygons of the original data set.
- 5) Graphic output from the computer duplicating each of the original data sets and the overlaid data sets was produced, as were reports listing the polygon code number, centroid coordinates, and descriptor data for each plot.

The commands to generate the above retrieval processes took 1.5 days to prepare, the elapsed time for all processes was 10 days,<sup>1</sup> and the total man-effort was 11.5 days.

1. The test reports were non-standard formats and the cost for generating the special-purpose reports is included in both the man-days of effort and the cost of computer use.

The data retrieval costs in terms of gross computer use are given in Table 5.6.

Linear measurements	\$ 25
Overlay	287
Sorting	35
Report program development	250
Report generation	57
Validation reports (DDSLT)	43
Total	\$697

Table 5.6. CGIS test: retrieval costs.

It can be seen that each of the retrieval processes involves high-order measurement and the extraction of data that are inherent in the file structures of graphic source material. As mentioned above, several of these processes are accomplished en route to the data bank. The overlay process, however, is not included in the formation of the data bank, and can only be carried out in CGIS after the data bank has been established. It is perhaps instructive to examine the cost of this process in direct comparison with manual techniques operating on graphic source data.

The overlay, which was carried out effectively, processed a total of 2,140 sq. in. of data. Assuming that 20% of the report-generating process was associated with the overlay (it was probably less than that, as overlay report is a routine CGIS function), then the total cost incurred was  $\$287 + \$70 = \$357$ . The elapsed time for the actual overlay run was less than 10 min, but, assuming that 2 days were nominally involved in request programming and report generation associated with it, the overlay cost \$357 and took 2 days.

From a manual overlay operation examined in detail in Chapter 3, pp. 75-76, with data sets of comparable density, the time required for 2,140 sq. in. of overlay, generating exactly the same reports, would require approximately 200 man-days (which, at a cost of \$5.00 per hour, would amount to \$8,000).

The total cost of completely creating the non-graphic data bank and performing all the retrieval procedures mentioned above was \$3,607.54, this figure including labour costs and gross computer utilization costs.<sup>1</sup>

From this limited test, it can be suggested that, given the need to perform

1. Overhead costs of physical plant, equipment other than the computer, and supervisory staff are not included in either the manual or machine-aided estimates. Overhead costs, including computer amortization, are included in the gross costs of computer use given.

high-order procedures for data extraction from a graphic source data base of irregular polygons, it is at least as efficient in terms of operating costs to create a non-graphic data bank of the CGIS type to carry out the operations concerned. In addition, if more than one retrieval (measurement or comparison) of data is contemplated from that data bank, then the comparable operating costs are markedly lower. The computer-aided techniques can also substantially reduce the elapsed time between the request for retrieval and the delivery of the results of high-order measurement and comparison of graphic data.

So far, this chapter has been dealing with the category of systems using an irregular polygon storage format, which can be considered to be under development in the early 1970s. The remaining part of this chapter will discuss current developments in input, manipulation, and output techniques which mainly reflect problems related to hardware. These encompass digitization, automatic line following, scanning, interactive manipulation of the data base, manipulation of remote sensing data, and graphic output capabilities. In each of these areas, developments in active progress will generally apply to all the categories of system discussed so far.

#### CURRENT HARDWARE-RELATED DEVELOPMENTS IN INPUT, MANIPULATION, AND OUTPUT CAPABILITIES

##### Input Techniques

The foregoing discussion has noted that if digital computers commonly available in 1974 are to be used to handle graphic images, then a non-graphic (digital) representation of the images must be generated. This image "digitization" has been recognized as a step with substantial time and costs related to it. As might be expected, several approaches to digitization are under development that attempt to reduce the cost burden and the inherent constraints on development of geographic information systems.

Current point/line digitizing. The simplest form of manual digitization is to place a transparent grid over the graphic image and to read by eye the row and column values of points. An equivalent form is the use of a set square or similar device to read coordinate values of a point from the coordinate

axis at the edge of a given sheet. Current semi-automatic "digitizers" are essentially electromechanical devices that use the same approaches as the manual techniques. A point may be electromechanically related to a grid underlying the surface of the digitizing table, or to the implied coordinate axis at the edge. The technique is thus essentially limited to that of a reading device, in that the location of the point or points is still chosen by the operator who positions a cursor. In most of such systems, a stream of points is generated as the cursor is moved along a line. Such digitizing units are reasonably well known and many types are in use, from those where the cursors are moved by hand wheels, to the "free pencil" electromechanical types, to the more recent Bendix datagrid. Another recent addition to the range uses a sonic position detector.<sup>1</sup> Hardware costs for current digitizers vary from \$5,000 to \$55,000. The available current equipment is fully discussed by Boyle in "Geographical Data Handling"<sup>2</sup> and will not be further described here. It is, however, pertinent to note the limitations of the current equipment, so that the potential advantages of digitizing systems under development can be understood.

Current systems necessarily rely on the tracing capability of humans. This is a low-order manual function, but it has been observed<sup>3</sup> that the maximum speed for tracing with good tolerance limits is only about 2 in. per min. This has obvious constraints on all line data input. The second difficulty is the possibility of error in assigning descriptor data to the graphic image. A third difficulty is, of course, the errors that may be introduced by the equipment itself. The fourth limitation is that the digitization process produces a "blind" record, that is, the operator cannot directly compare the record he is producing with the source document, and thus errors can only be detected by subsequent plotting and editing of the result. The limitations of current equipment are graphically illustrated by the following extract from the PIOS report.<sup>4</sup>

" The Geography Department... maintained a sub-contract... for digitizing, pre-editing and re-digitizing of the polygon data. Hardware and other problems experienced by the University seriously complicated the project situation and more than doubled the estimated time required

1. The "graf/pen" produced by Science Accessories Corp., 65 Station St., Southport, Conn., U. S. A. 2. Tomlinson, R. F. 1972. 3. See tracing capability tests detailed in Ch. 3, pp. 71-73, and discussion of digitizing operator performance by Boyle in Tomlinson, R. F. 1972. 4. Dangermond 1971.

for digitizing the maps.

" Originally the University digitizing equipment being used was configured for punched card output only, and had been subject to only light to moderate workloads. By the end of the first month of heavy digitizing, sporadic equipment anomalies had introduced enormous editing problems. A magnetic tape recorder was requested and installed.

" Once operational, the tape recorded output showed poorer quality than the punched cards. After extensive investigation it was found that similar problems experienced by other users of this equipment had been traced to service voltage variation. Therefore, the various units were isolated on separate circuits and voltage regulators introduced. These measures corrected punched card errors, but had little effect on magnetic tape output. Next the magnetic tape and interface was investigated. By insulating the unit's chassis, providing an absolute ground, shielding cables, and careful tuning, the most severe kinds of errors were eliminated and error frequency substantially reduced. One week of production showed a rate which, though much lower than original estimates, could be ameliorated with increased staffing. This progress, coupled with the favourable appearance of the first plotted maps, caused... the University to proceed.

" The fourth and fifth project months showed reasonable progress. Noise appeared in output tapes periodically, but frequent retuning of the recorder interface worked to alleviate the problem. In addition, a special computer program was written to overcome the impact of noise on processing. An unexpected further problem appeared in that digitizer operators began to exhibit an ability to "overdrive" the recorder. This "overdrive" problem was caused by operators who attempted to record a greater number of points than the machine had the capability of storing and writing on tape. The digitizer also showed some cumulative drift, but not of a significant amount when the origin was regularly checked and reset. It was thought that the operators would self-discipline their activities. In many occasions this discipline did not prevail and the job of redigitizing increased as a result. University staffing did not reach expected levels, but increased steadily.

" The second batch of maps were plotted during the last part of the fifth project month and the first part of the sixth project month. They illustrated that remaining sources of errors were more serious than anticipated. One of the student digitizer operators was producing work of poor quality and in many cases all the operators were producing drift. Several of the other operators continued to overdrive the system. Computer programs were modified to help correct drift and a rigid quality control plan was designed but much redigitizing was needed. Toward the end of the sixth month the digitizing equipment suffered breakdowns. This breakdown necessitated extensive repairs.

" During the seventh and eighth months of digitizing it was found that the final digitizer repairs, including a field modification to length circuit time constant in the tape interface, almost totally eliminated noise and sharply reduced overdriving. It was demonstrated that if the equipment was operating effectively to begin with the delay and error problems would have been greatly reduced."



It should be emphasized that not all digitizing efforts have the same problems, and that the problems arise mainly when the volume of data necessitates a production operation rather than an experimental operation. This author, however, doubts that any current production operation involving digitizers has reached any level of efficiency without experiencing similar difficulties, even if such difficulties have not been reported.

Interactive point/line digitizing. Problems similar to the ones noted above prompted the use of computers to monitor the record being produced by the operator during digitizing. The availability in 1968 and more recently of mini-computers that could be economically dedicated to a digitizing process, coupled with accurate, inexpensive, on-line display units using cathode ray tubes (CRT), greatly accentuated this trend.

In this interactive process, the computer checks that the data are properly recorded, that the descriptor formats used by the operator are sensible, that the "line data" sensibly represent a line of desired resolution and accuracy, and that "island" or "polygon" closures are completed properly; it also allows the operator to backtrack and recover or correct errors as they are made. The interactive digitizing approach is still limited to the tracing speed of the operator, but greatly improves the quality of the records produced and goes far toward eliminating the lengthy procedures of editing and error correction necessitated by "blind" digitizing.

Several such interactive digitizing systems are under development. Well-known ones are resident in the Royal College of Art,<sup>1</sup> London; Rome Air Development Center,<sup>2</sup> U. S. A.; United States Geological Survey,<sup>3</sup> Washington, D. C.; and the Canadian Hydrographic System<sup>4</sup> developed at the University of Saskatchewan, Canada. This approach definitely has merit, particularly when the work requirement involves low to medium volumes of data. Numerous firms in North America, including IBM, Raytheon, H. Dell Foster, Synercom Technology, and others, have adopted the approach for systems currently under development and still proprietary. The hardware cost for such systems varies from \$50,000 to \$150,000. In each case the software has been developed by the agency concerned and has not been implemented elsewhere.

1. Bickmore 1966, 1971, Rhind 1971. 2. Diello 1970, P. R. C. 1973.  
3. U. S. G. S. 1969. 4. Boyle 1970a, 1971.

An interesting hybrid version of this approach is used in the COPIER system<sup>1</sup> (Lockheed), where a graphic image is optically projected on to a CRT surface and a light pen is used to trace the image interactively. Somewhat surprisingly, a grid block structure is derived from the COPIER system for ease of subsequent processing in the relatively small areas that have been attempted, though the Lockheed file structure described has quite adequate boundary-handling algorithms.

Automatic line following. If a device could be developed that automatically follows a line (and accurately records the path followed) at a faster rate than a human can guide a cursor, then a significant constraint on line digitizing would be removed. This is a goal that has been sought but not yet won. In practice, the approach used concentrates on the procedures whereby the operator predefines a line, manually traces any "lengths of confusion", and allows the line follower to fill in the detailed line data.

The Gerber Company, Mass., U. S. A. and the U. S. G. S.,<sup>2</sup> Washington, D. C., independently have done considerable research on this approach, but have not yet developed an acceptable system. The University of Saskatchewan<sup>3</sup> uses a Vidicon camera carried on an X, Y mechanism, connected to a PDP8/E computer. The Vidicon tracks in steps of about 1/2 in. along each line in turn. Within each step a scan operation is carried out, but the final output is in line form and is directly comparable with output from other line digitizers. In order to make the line tracing operation completely automatic, the work is preceded by a visual examination of the area on the interactive display CRT. Start coordinates and descriptors are then allocated by the operator. Speeds of approximately 60 in. per min can be achieved using the operations of pre-editing plus automatic line following. This approach can be said to be under active development. Costs of hardware associated with it vary from \$30,000 to \$200,000. No automatic line following system is in productive use in any geographic information in 1974. However, the generation of line information in 20 to 30 times less time and at measurably lower cost than current digitizing systems is a probable future development. The usefulness of this approach will depend upon the speed with which it can be brought into

1. Merrill 1973. 2. Personal communication, Edson 1973. 3. Brown 1970.

reliable production, the architecture of large-scale computers, and their needs and efficiency in handling line data formats at that time. Given the present architecture, capability, and file structures of computers, the automatic line following approach would seem to offer extremely attractive advantages.

Scanning. The scanning approach essentially uses some device to record a patch (or all) of the surface of the source document in machine-readable form, and relies on the computer subsequently to identify the image elements within the patch that are of interest to the user and need to be retrieved. The approach requires unequivocal data to be scanned. The programs needed for subsequent identification and labelling of line segments are sophisticated; in addition, a manual or semi-automatic process must be used to relate descriptor data to the images.

Two versions of the scanning approach have been followed. The first is based on the well-known facsimile principle, where the source document is placed on a drum and rotated in front of the scanning head, which moves incrementally across the face of the drum. The entire document is scanned and translated into bits of data, the number of bits for any one document depending on the scan spot size. In the second version, a small area (about 2 in. x 2 in.) of the source document (or a microfilm of a larger document) is scanned with a flying spot (raster) scanner, and the complete data for that area are recorded on magnetic tape. This latter process takes seconds only. The data are subsequently handled in much the same way as those from a drum scanner.

The raster scanning approach would seem to be of immediate benefit. However, the design problems have not yet been overcome.<sup>1</sup> A flying spot scan of 0.001 in. resolution results in 1 million spots per sq. in., or 0.6 billion spots for a 20-in. x 30-in. map. A 0.01-in. spot size gives only 10,000 spots per sq. in. or 6 million spots for the same map. Therefore, spot size, shape, spacing, precision placement, overlap, energy distribution, and other parameters of spot design must be given extremely careful consideration to provide the system with economic scanning, plotting, processing, and display. Processing time per map can quickly become

1. See discussion by Sharp, pp. 687-689 in Tomlinson, R. F. 1972.

excessive with the computers in general use in 1974.

The drum scanning approach, though slower (between 1 and 15 min, depending on map size) has been used to digitize economically source documents of up to 45 in. x 45 in. in size.<sup>1</sup> A Visicon drum scanner with a resolution of 0.01 in. is used for documents of up to 11 in. x 17 in. at Pennsylvania State University.<sup>2</sup> The Rome Air Development Center<sup>3</sup> uses a large (45-in. x 45-in.) precision drum scanner, as does TOPOCOM (U.S. Army Topographic Command) in Washington, D.C. The RADC scanners are intended for both plotting and digitizing (in that order), and are part of the proposed system of data storage in graphic form. The TOPOCOM unit was primarily intended for digitization, but uses an early form of display manipulation system to aid the software in joining small areas recorded from the scan operation. The Canada Geographic Information System uses a drum scanner, as described in the account of the CGIS given above and in detail in the Appendix. The input procedures, descriptor data linkage, and subsequent processing steps given in the CGIS description are typical of those encountered in the drum scanning approach. The equipment costs for drum scanning vary from \$50,000 to \$180,000.

An interesting hybrid approach<sup>4</sup> is used by Calspan Corporation (formerly the Cornell Aeronautical Laboratory, C. A. L.). The initial step for digitizing bathymetric charts semi-automatically is to scan the chart and produce a digital image, which is displayed to the operator on a CRT. After geometric referencing of the edge "tick" marks, a set of closely spaced parallel (track) lines is automatically generated over the image. The chart is then rescanned along the track lines, and intersections with the bathymetric contours are automatically detected. Next, the information on depth for the intersections is automatically added if enough data are available. The operator may then interact with the system to detect and add missing contours, detect and delete extraneous contours, and provide the necessary additional data on contour depth. Track line by track line, the identified information is written into magnetic tape storage. This type of approach is of considerable interest, as it combines the swift aspects of scanning with the interactive capability of the operator to produce a form of automatic

1. See CGIS cost summary in Table 5, 3, p.162, and in Appendix. 2. Personal communication, Williams, C. M. 1973. 3. Diello 1970. 4. Solosko 1973.

line following. Costs of using the Calspan system are not available at this time. Such systems must be considered to be under development.

It can be seen from the above that the inherent constraint of graphic to non-graphic conversion is recognized not only from a theoretical standpoint, but as a problem that must be overcome if spatial data systems are to develop further. In technical terms, it is certainly the major constraint on system development at present, although it would appear that a significant reduction in the constraint might be reasonably anticipated within the next three years.

#### Data Base Manipulation Techniques

Interactive manipulation of spatial data base. A significant difference between graphic and non-graphic forms of spatial data storage is that the former are visible to a human and the latter are visible only to a computer.

With graphic data storage, the storage medium is the display medium, so data retrieval can begin instantly. Browsing is greatly aided by the random access to map sheets. When simple indexes allow identification of limited search areas and the available data sets are conveniently grouped, the manual process is rapid, efficient, and inexpensive. In contrast, in any of the systems employing non-graphic forms of data storage described in the previous chapters, the computer must be carefully instructed to carry out the simplest of reading tasks, browsing is completely inefficient (involving, as it does, the requirement to specify what must be browsed before the browsing begins), and when all is done the computer output (in graphic form) still has to be read by a human.

With these comments in mind, it is interesting to examine systems under development that both allow direct visual examination of non-graphic stores of spatial data via a cathode ray tube, and offer the ability to browse through the data base, select sets of data of interest to the observer and, through the same medium (CRT), perform (direct and receive results from) some manipulation tasks on the selected data.

Although this interactive interrogation of a data base is not uncommon in

large non-spatial data sets,<sup>1</sup> it has not been widely adopted for large-scale manipulation of spatial data.<sup>2</sup> Perhaps the best example of spatial data systems that employ the approach is the appropriately named INSIGHT<sup>3</sup> (Interactive System for Investigation by Graphics of Hydrological Trends), a system developed by R. F. Phillips of Unidata, Inc., Ann Arbor, Michigan.

The INSIGHT system uses a storage tube terminal (Teconrux 4010) connected to a time-sharing computer (IBM 360/67)<sup>4</sup> to display the locations of a point data set of water quality monitoring stations, superimposed on a background of geographic features. Through the CRT, the user can manipulate this map background in a variety of ways and ask for and receive tabular and graphic presentation of the water quality parameters that have been measured for any of the point data stations.

The cartographic data base is quite small and essentially fixed (approximately  $2.2 \times 10^6$  bytes of data). The data base related to the point data set of water quality monitoring stations is large (currently  $700 \times 10^6$  bytes) and is expandable.

Since the system is geographically oriented, the user's first step is to select a reference map from the cartographic data base. The State is the basic cartographic unit of the system. The user can have any state map displayed, or any group, or all of them, and then indicate the one or the group that is of further interest. The resulting map can be modified, by shrinking or expanding the display scale, by zooming or panning, with respect to the original data, or by designating a certain polygonal portion (a county, for example). Additional geographical features such as county boundaries, river traces, lake outlines, and municipal boundaries can be selectively added at this point by requesting that they be overlaid on the state outlines. This is essentially a browsing process through a set of maps. The scale can be enlarged until the limit of the original digitized data is reached.

1. As, for example, at airline ticket counters in North America and Europe.  
 2. It has, however, been adopted for a variety of military applications, and in commerce, for on-line manipulation of spatial data for air traffic control.  
 3. Phillips, R. L. 1972, Unidata, Inc. 1973. 4. This is a duplex cpu configuration with 2,700 pages of virtual memory (4K bytes/page) available on three paging drums. The communications interface is a data concentrator based on a PDP8/E and associated multiplexing and port selection hardware. Conventional support routines with teletype devices are used for graphic terminal operations.

Several additional manipulation features are possible on this data base. Areas can be measured by tracing their outline with the cursor on the CRT. Similarly, lines can be measured by moving the cursor along their length. Longitude and latitude of any cursor position can be calculated. The results of these calculations are shown on the screen. The most important facility of the system is, however, the connection with the hydrographic data. The cartographic base acts as an index to the vast store of water quality data.

The point data representing the positions of the water quality stations are displayed on the cartographic data base. From this base, the desired set of stations is selected. A list of data available at each of the selected stations can then be displayed. After browsing through the list, the observer selects the desired variables. A conventional graph showing the distribution of the variables over time at the selected stations can then be generated; equally, the data can be listed in tabular fashion or, perhaps more interestingly, the water data and cartographic data base can be combined at this point and a three-dimensional perspective display of the stations, the country surrounding them, and the measurements of the specified parameters can be displayed as a vertical bar graph.

General cost figures are not available, but data have been gathered from operation in the University of Michigan computer. When that system is used, one is charged mainly for: (1) connect time, at \$3 per hour; (2) cpu time, at \$4.73 per min; and (3) actual virtual memory, at \$2.85 per page-hour. For a typical 1-hour session with INSIGHT, and 20 to 30 retrievals for about 20 stations, the cost runs between \$15 and \$20.

The limitations of the system are obvious. It is a point data system of the GRDSR type, embedded in a simplistic cartographic data set. The simplicity of the cartographic data set is ideal for the indexing purposes for which it is used, and no precision is attempted or needed. The spatial manipulation functions (area measurement and coordinate point measurement) are essentially manually traced, and the operator works in a relatively expensive (computer-linked) environment. Nevertheless, the ability to read effectively and browse through non-graphic storage at these volume levels is clearly indicated, as is the ability to gain rapid access to huge

amounts of data and manipulate selected subsets. This process would seem to be a fundamental requirement if large-scale non-graphic stores of spatial data are to be of wide use in the future.

Manipulation of ERTS numerically coded images. Remote sensing is a survey technique that can and does produce large quantities of spatial data in non-graphic form as a primary product. The ERTS-A satellite, for example, produces approximately 9,212 original numerical images of 1,316 scenes over the U. S. A. alone, each week. Of these, between 1% and 5% are retained in numerical form. Each such image, however, contains approximately 16 million cells, each coded to 64 grey levels in 6 bits; in other words, each image produces 96 million bits of numerical information. Each week approximately 300 numerical images are recorded on approximately 870 computer-compatible magnetic tapes. The original image data produced can, of course, be translated into graphic form by a variety of processes. The resulting pictures can be visually interpreted and the results of such interpretation can be converted to data bank formats using any of the techniques described above.

A considerable challenge, however, exists in the possibility of using the computer to manipulate and convert directly the 1% to 5% of images left in the raw numerical data format, into a useful data bank format. The many variables associated with this task are thoroughly documented by Steiner in "Geographical Data Handling".<sup>1</sup> The data reduction steps include image processing, pattern recognition, statistical relationships to non-observables, descriptive unit assignment, change detection, and geometric base reconstruction. The resulting products would be amenable to encoding for data bank use or to a future form of pictorial information system.

It is outside the scope of this thesis to explore the data techniques involved in pictorial data reduction. They do, however, represent an extension of the conventional non-graphic processing techniques covered in this thesis and will form a bridge between the large masses of numerically coded picture information now becoming available and geographic information systems that will be developed hereafter. At the moment, however, the techniques

1. Tomlinson, R. F. 1972.



devised for numerical data reduction do not allow accurate thematic information to be extracted from the numerical pictorial data base, except in closely controlled test situations or for gross thematic observations. In addition, the geometric restitution of the pictorial images degrades their already low resolution and further invalidates thematic identification processes.

In the main, it can be said that this approach to manipulation of raw numerical spatial data does not add any significant amount of data to geographic information systems, but that it is an approach which must be followed further as sensors are designed to produce more precise, high-resolution data and as computer configurations are developed that are amenable to such processing.

#### Output Devices

Output devices are those that convert non-graphic data to a graphic format. They range from line printers of alphanumeric symbols (or special symbols), through symbol and line drafting machines (plotters), to cathode ray tubes of various kinds. There is a wide range of designs and capabilities, costs and conveniences in these machines. A full description of the equipment has been provided by R. Boyle<sup>1</sup> in "Geographical Data Handling". The spectrum of capabilities for producing maps with these devices was thoroughly demonstrated by D. Edson<sup>1</sup> in the same volume. Their reviews will not be repeated here; it is simply pertinent to note that the devices present no constraint on current ability to handle spatial data. Any non-graphic image can be readily and economically transformed to a graphic image, to any degree of precision and resolution inherent in the non-graphic data. Each year such systems develop increased speed of operation, improved reliability, and lower unit cost. This, at least, is the situation in 1974. Perhaps with the exception of CRT storage tubes (which might conveniently be somewhat larger), no spatial data base has graphic formats that cannot be displayed within the limits of existing devices. One can, however, visualize future developments in spatial data storage (such as holographic memories, or picture-processing computers) that would exceed the capability of existing output hardware. This stage has not yet been reached, and it is reasonable to say that the current output equipment and the techniques employed are as good as or better than current capability for spatial data handling.

1. Tomlinson, R. F. 1972.

### MULTI-FORMAT SYSTEMS

A multi-format geographic information system is one that attempts to handle more than one format of non-graphic spatial data. There are two aspects to this capability, either or both of which may be present in a multi-format system. The first is the ability to accept and store data of different formats in the basic file structure of the system. This process is clearly different from that of converting different source data formats to a single format for subsequent non-graphic storage. The second aspect is the ability to change the data stored in single or multiple formats in the file structure of the system to alternative formats for use in subsequent procedures of data manipulation or display. These two aspects will be considered separately.

The ability to change data stored in a single format to alternative formats is increasingly common in geographic information systems. The purposes behind such transformations are varied. A typical use is to reduce the volume of data that must be handled in subsequent data manipulations, or simply to provide the data in a format that is amenable to certain types of manipulation. Another use is to allow comparison with data sets that are available in the alternative format. As with the transition from the format of real-world data elements to the storage format, a change may imply some loss of information. Similarly, the loss of information tends to decrease as the size of the storage element decreases in relation to the size of the output elements, as shown in Fig. 5.15.

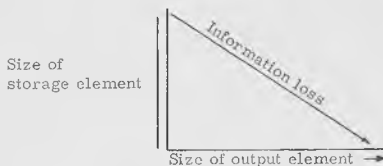


Fig. 5.15. Information loss in relation to sizes of storage and output elements.

Very small grid cells, for example, may be grouped into very large irregular polygons with a minimal information loss (and with an attendant reduction in data volume). A polygon represented by its centroid coordinate, on the other

hand, may incur a substantial information loss (although the information loss may not be of concern to a particular user, and the significant reduction in data volume for some manipulations may be attractive). The following diagram (Fig. 5.16.), which closely follows Fig. 4.7 but which now relates storage format to output format, shows typical transformations which are currently carried out in various systems.

		Storage format				
		GRID		POINT	LINE	POLYGON
		Small	Large			
Output format (for manipulation or display)	GRID	Optimum		GRDSR ✓		CGIS ✓
	POINT	Same	Same	Optimum		PIOS ✓
	LINE	SYMAP ✓		DIME, FRIS Arizona ✓	Optimum	
	POLYGON	SYMAP ✓		GRDSR ✓	DIME ✓	Optimum

Fig. 5.16. Transformations between storage and output formats undertaken by typical systems.

The advantages of handling reduced volumes of data (for example, polygons represented as centroids) can be readily understood. Similarly, the desirability (if not the validity) of changing formats to allow comparison with other data sets can be appreciated. The value of changing non-graphic storage formats to facilitate specific spatial data manipulations is illustrated in the following examples.

Given a surface described by data elements in the form of irregular polygons whose boundaries are isolines, and given the need to determine surface slopes, it is theoretically possible to calculate the slope data directly from the isolines, but only with a relatively complex algorithm which is thus expensive to use. The polygon format can, however, be changed to a small grid format from which slope determination is simply derived by comparing the elevation at a given cell with that of the surrounding cells. Division by the

constant scaled distance between rows and columns gives the slope gradient. The cell size chosen is, of course, significant, as the calculation cannot take into account irregularities that may exist within the grid cell. If the cells are relatively small, for the majority of decision-making purposes the results obtained by this approach to slope calculation are directly comparable to results from the isoline interpretation algorithm, and they are substantially easier to acquire.

Similarly, values of irregularly distributed point data can be used to calculate related values for an overlay of small grid squares (as opposed to grouping points within grid squares). The resulting array can be used in data manipulations such as diffusion analysis with much greater efficiency than direct use of the point data set (though error may be introduced by the assumptions inherent in the point-to-grid transformation process, and may arise from the original distribution of the points).

The examples given above might indicate that a grid format is a desirable primary format for data. In some circumstances it is. It is aligned to much computer architecture (particularly that of the early computers), and even to some computer languages (FORTRAN, for example, has its roots in matrix methodology). However, it must be stressed that the format change is, to a great extent, irreversible. In the example given above, it is not possible to recalculate the locations of the original point data, or to trace the original polygonal boundaries accurately from the grid values. The ability to change to a grid format for some manipulations is clearly desirable. Its value as a unimodal storage format has to be considered with due caution and with attention to overall requirements for data manipulation.

In the same manner that point and polygon data can be related to grids, so data from alternative formats can be related to line segments. Just as the example given above of transforming point data to grids implied the assignment of a weighted value to a grid cell, so can the distance-of-point data attributes be used as a weighting factor in assigning values to line segment data. This approach can, for example, internally simplify the analysis of corridors through line networks. A recent study to locate transmission

lines carried out for the Arizona Public Service<sup>1</sup> weighted data items for relative compatibility with transmission line routes, considering such factors as distance from historic sites (point data), distance over bad ground conditions (polygon data), and intersections with railroads (other line segment data). The study derived values for each line segment based on this transfer of value from one data format to another. The resulting analysis for network optimization was a straightforward linear procedure.

Just as in the case of grid cells, allocation of the (weighted) values to the line segments is irreversible. It is not possible to recreate the source data format from the new format. However, the benefit of the capability for format change within systems can be appreciated.

As was noted in Chapter 4, pp. 96-97, at this early stage of development of computer-aided systems for handling spatial data there has been a tendency to avoid the complications of cross-format linkages that would be required by multi-format primary storage; hence, the file structures of systems have generally been developed around a single format of spatial data. The pressures to change this state of affairs are easy to understand. The real world can be more accurately described if a mixture of data formats (points, lines, and areas) is employed, and explanations of the interaction of real-world phenomena frequently involve real-world data elements that have different optimal storage formats. The design of several systems obviously reflects a response to these pressures.

The FRIS system from Sweden, now being further developed by the NORDPLAN<sup>2</sup> agency, has altered its primary point file structure so that line networks can be identified and manipulated. The point data approach is still used to locate the basic statistical records. The point-network linkages have allowed the development of quite sophisticated models for resource allocation. Simple fixed-source districting problems, such as allocation of school boundaries and travel-to-work calculations, are now being carried out in addition to manipulations of point data.

The DIME<sup>3</sup> system, although primarily an example of a line network system, is also an example of a multi-format system that accepts areal and line data

1. Dangermond 1973. 2. Personal communication, Salomonsson 1973.  
3. Corbett and Farnsworth 1971.

into its primary storage format (hence the word "Dual" in its acronym). The system stresses its capability for handling line networks and has some inefficiencies in terms of describing complex areas, but nevertheless it handles both formats.

San Diego County<sup>1</sup> in California is in the process of combining the file structures of its coarse polygon system (a version of PIOS) and the DIME system. The degree to which this has been successful is not known.

The CGIS now has the ability to accept data related to point formats. This is a fairly recent development<sup>2</sup> (1973). It would seem to lead the way to linking the capabilities of CGIS with the file sophistication of GRDSR. As both systems have been developed within the same government, this may present an opportunity for complementary development. It is, however, significant to note that the CGIS has not attempted (and at the time of writing has no plans to attempt) to acquire the ability to store network data in its main file structure.

Undoubtedly, the ability to store various data formats in the primary file structure of a system is being considered, and development along these lines is taking place. Equally, such an ability would be desirable. Consideration of the file structures of the typical system described in this thesis, however, does not lead to the conclusion that modification or merging of existing designs will be an easy process or will lead to efficient processing. It is probable that efficient file structures for multi-modal input will have to be designed with the mixture of data formats explicitly predetermined.

#### SUMMARY

This chapter has focussed on the status of techniques under development for the non-graphic storage and computer-aided manipulation of spatial data. The irregular polygon category of systems for handling spatial data was examined and its characteristics illustrated with specific examples. The computer-aided processes were compared with equivalent manual processes. The trends in the development of digitization techniques, interactive manipulation of the data base, output devices, and multi-format systems were also examined, and their potential contribution to processes for handling

1. Personal communication, Taylor 1973. 2. Personal communication, Switzer, W. 1973.

spatial data was considered. Several conclusions can be drawn from the evidence presented.

The techniques of graphic to non-graphic conversion that were noted to be major constraints on current systems remain as constraints on systems under development. Dangermond's<sup>1</sup> graphic description of the difficulties of digitization encountered in the relatively advanced PIOS system was presented. Only when more advanced digitization techniques are employed, such as use of the drum scanner by CGIS, are the overall costs of computer techniques lowered to acceptable levels in comparison with those of even medium- to high-level manual processes. This comparison was clearly shown in the CGIS cost analysis. The constraints, however, are being removed. Speeds of digitizing 20 to 30 times faster than current methods can be reasonably expected. This, coupled with the advent of substantial volumes of numerical data directly derived from survey processes (such as ERTS-A data), can be predicted to result in greater volumes of metrically coded non-graphic spatial data in the foreseeable future.

The prospect of increasing volumes of non-graphic data, both by improved digitizing or different gathering techniques and by the use of high-order data formats such as irregular polygons (or low-order, high-volume data formats such as grids), brings with it the problem that the processing capacities of computers may be exceeded. Already (in CGIS) it can be noted that some processing techniques (sectioning) were determined by computer processing capacity and that the file structure (frames) was determined by anticipated data volumes (overlays) related to computer processing capacity. This is not a function of storage capacity of a computer (which is practically infinite), but of the access time to the various forms of storage. As present computers are apparently approaching the theoretical limits of electro-mechanical storage, a new approach to computer memory (optical storage, holographic memories) may have to be implemented to remove this constraint. This factor does not inhibit the use of current systems or seriously influence those under development, but it will have significant impact if non-graphic data volumes are increased, and there is every indication that they will be.

1. See quotation on pp. 169-170 in the text above from Dangermond 1971.

In contrast to the limitations imposed by the transformation of data to non-graphic formats and the potential constraints of computer processing speed, there is a significant under-use of the architecture of current computers by even the most advanced spatial data handling systems. It is not necessarily true that the most complex file structure is the most efficient, but the file structures employed by the systems under development are extremely straightforward. These considerations lead to the general observation that the great majority of computer-aided systems in any state of development have been designed in the absence of requirements to manipulate data in a sophisticated manner. The only systems that are designed to operate in a user environment, under the control of a system monitor and with a suite of commands to generate data manipulations, are the CGIS and INSIGHT. Neither of these systems could be considered to have used the full capacity of present-day computers for their file structures. The suite of CGIS commands is limited, in any sense of interrogative spatial analysis, and the INSIGHT file structure facilitates reading and browsing rather than any substantive manipulation. Neither is linked to any form of spatial analysis program or set of procedures that require complex data manipulations. It is clear that future system development will need greater interaction and direction from system users. As spatial data volumes increase, and if these data are to be manipulated by computer, an increase can be expected in the requirements for spatial analysis of the data in a non-visual manner and for combined visual and non-visual (interactive) analysis. Certainly, if improvements are to be made in systems under development, they can most beneficially be applied to these processes. The relationship between data availability, data handling capability, and user requirements will be examined in the following chapters.



CHAPTER 6 - ANALYSIS OF SPATIAL DATA HANDLING CAPABILITY  
IN TERMS OF DATA CHARACTER AND AVAILABILITY:  
A FRAMEWORK FOR DESCRIPTION OF GEOGRAPHIC  
INFORMATION SYSTEMS

The two previous chapters have examined the status of existing computer-aided techniques for handling spatial data. The purpose of this chapter is to identify the contribution of the various techniques to our overall ability to handle spatial data. A tentative framework is provided within which spatial data handling systems<sup>1</sup> can be located and their relative capabilities understood. By extension, it can also be used to introduce the relationships between handling capability and the requirements of the user.

ANALYSIS OF FRAMEWORK

From the standpoint of these systems, spatial data can be characterized by the nature of the location identifier used, and the volume (number of pieces) of data carried in concert with any one set of location identifiers. Location identifiers range from the most elementary form of external indexes to explicit boundary definition on the surface of the earth, with a positional accuracy of plus or minus half the width of the line. The volume of data varies in scale from one or two pieces of data per location identifier to large amounts up to 50,000 pieces of data per location identifier, arranged in overlapping categories. The third basic variable of a spatial data handling system is the range of capabilities for manipulation; it extends from elementary forms of retrieval up to complex data handling operations under the control of a system monitor.

These three variables (location identifier, data volume, and manipulation facility) can be thought of as three axes of a cube (Fig. 6.1) within which the systems described in Chapters 4 and 5 can be located. The use of the diagonal axis of the cube as a vector of potential file-structure capability will be examined later in the chapter.

It must be emphasized that the values of the variables along each axis are essentially arbitrary, and their relative positions are useful only as general indicators of system characteristics. The progression along each axis is certainly not linear, and in at least two cases tends to be logarithmic, as

1. A "spatial data handling system" in this sense is taken to be a set of one or more computer-aided techniques that can read, store, manipulate, and display spatial data.

will be shown. There are discontinuities in the series of values and, hence, implicit steps in the scales. In short, the internal space of the cube is not simply defined, and systems cannot be precisely located within it. As more information becomes available about various systems, it may be possible and worthwhile to regroup or refine the values. The approach is provided here only as a convenient framework for discussion of system characteristics.

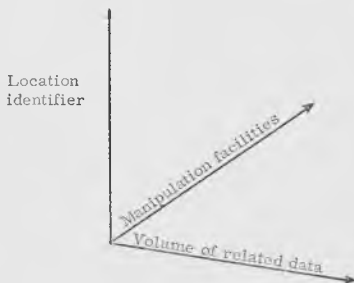


Fig. 6.1. Axes of descriptive framework.

The range of values along each axis will be examined briefly below. Thereafter, the various systems will be located with reference to the axes and their relative characteristics will be discussed.

#### Location Identifiers

The various approaches used to define and code the location of spatial data were discussed in some detail at the beginning of Chapter 4. To summarize and simplify that discussion, there are essentially three categories of methods for identifying location. Arranged in ascending order of resolution, the first category gives a name to the location of the data element without otherwise defining its position, the second describes the relative location of data elements, and the third, using various levels of precision, defines the location of the boundary<sup>1</sup> of the data element. The series of location identifiers, with examples of their application, is illustrated in Fig. 6.2.

1. All real-world elements of spatial data have boundaries. Points are regarded as data elements with boundaries but no significant dimensions, lines as data elements with boundaries but no significant dimension in one direction, and areas as bounded data elements with two dimensions.

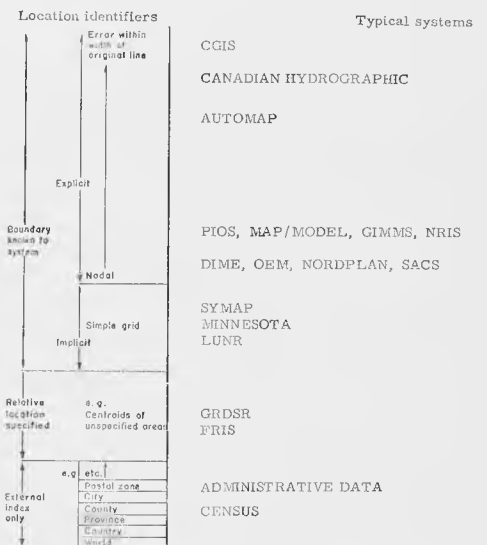


Fig. 6.2. Location identifiers with related data sets.

The simplest category of location identifiers is the external index, a nominal code that identifies data as belonging to a particular geographic area or location. The value does not convey relative location, which in this case must be determined by reference to an external metric index, typically a map. Within this category it is possible to recognize a hierarchy of names referring to decreasing sizes of data element (and hence increasing resolution of location), for example: world, country, province, county, postal zone, street name, and street address. Probably the overwhelming proportion of non-mapped data with any locational identifier today uses this descriptive method. Most traditional census and administrative data, for example, are coded only by this method.

The second category consists of identifiers that convey the relative location

of the data elements, but do not define the position of their boundary. A typical example of this category is point coordinates that represent the location of otherwise unspecified areas, as in GRDSR<sup>1</sup> and FRIS.<sup>2</sup> Another example is the DIME<sup>3</sup> system, which provides a topologically rigorous connectivity code for city streets without necessarily mapping the resulting network in space.

This last example illustrates one of the difficulties in absolute allocation of systems to categories. In the DIME system, coordinate values can be added to its network nodes, and hence it can be regarded as a higher-order system (see Fig. 6.2). Also, several higher-order systems have the ability to change the data format. Such changes are frequently made to lower-order location identifiers, and when operating in that mode the systems might be considered to be of a lower order. In general, however, systems are allocated to the category that represents the highest function they can carry out.

The third category of location identifiers comprises those used in systems where the position of the boundary of the data element is implicitly or explicitly known.<sup>4</sup> The lowest order within this category uses cells in a simple grid whose origin is metrically defined. The resulting locational codes may be metric (cell centroid coordinates, for example) or topological (such as cell sequence numbers). However, the essential characteristic is that the location of the actual boundary of the cell is (implicitly) known to the system, without being further (explicitly) defined by all corner points or by a string of coordinates of the boundary lines. There is a fine distinction in locational precision between use of centroid coordinates of undefined areas, as in GRDSR, and the grouping of data elements of otherwise unspecified location in large grid cells, as in LUNR.<sup>5</sup> The implicit definition of a boundary between sets of data by the grid cell method does, however, place it in the lowest order of the third category. Within this grid cell sub-category there appears to exist a hierarchy of values, ranging from large to small cell size (in relation to the size and regularity of distribution of source data elements), that is, from Type 1 arbitrary regular-area systems<sup>6</sup> such as LUNR to Type 2 arbitrary regular-area systems<sup>6</sup> such as SYMAP.<sup>7</sup>

1. Statistics Canada 1972a, 1972b; see also Chapter 4, pp.111-115. 2. Alfredsson et al. 1970, Melander et al. 1971. 3. Corbett and Farnsworth 1971; see also Chapter 4, pp.116-122. 4. "Known" in this sense means that the boundary properties concerned are recorded in the storage of the data handling system and are thus known to the system. 5. Hardy et al. 1971a; see also Chapter 4, pp. 99-103. 6. As defined in Chapter 4, p.98. 7. Harvard University 1968; see also Chapter 4, pp. 105-109.

Explicit boundaries of data elements form the higher orders of the category. Lines described by widely spaced nodal points may be thought of as simple (lower-order) examples of explicit boundaries. The hierarchy extends upwards as the resolution of the image description increases, as, for example, when the number of points used to describe a line increases. Because location identifiers can be used for various types of data element (points, lines, areas) within one system, it is difficult to allocate systems absolutely to categories, but allocation of systems is generally based on the level of location identifier used to define its most complex data element. Typical of the simplest form of explicit boundary identifiers are those used in the DIME, OEM,<sup>1</sup> NORDPLAN,<sup>2</sup> and SACS<sup>3</sup> network systems. Higher in locational precision but still at the "coarse polygon" level are the GIMMS,<sup>4</sup> PIOS,<sup>5</sup> MAP/MODEL,<sup>6</sup> and NRIS<sup>7</sup> systems. AUTOMAP<sup>8</sup> and the Canadian Hydrographic System<sup>9</sup> attempt to define their boundaries with greater precision. Typical of the higher-order systems is CGHS,<sup>10</sup> which locates all boundaries with an error of less than half the width of the original line.

The spacing of the categories and subcategories along the axis illustrated in Fig. 6.2 is completely subjective, but it is straightforward, and recognizes the ascending order of resolution of locational identification in the different categories. The lower half of the axis is occupied by techniques that do not explicitly define the position of the real-world data element, but use some surrogate.<sup>11</sup> This lower part of the axis is arbitrarily subdivided into three equal parts, each representing a more precise method within this approach to location identification; within each subdivision, an ascending order of locational resolution is assumed. The upper half of the axis is occupied by techniques that make an attempt to define the position of the boundaries of the real-world data elements in space. Again, an ascending order of locational resolution is assumed within the entire subcategory. The relative resolution of location identifiers used in the different systems can be identified reasonably clearly from this arrangement along the axis.

1. Observatoire Économique Méditerranéen 1971. 2. Personal communication, Salomonsson 1973. 3. Calkins 1967, Barb 1968. 4. Waugh 1972. 5. Dangermond 1971; see also Chapter 5, pp. 145-152. 6. Arms 1967, 1968, 1969, 1970. 7. Boeing 1972a, 1972b. 8. Schmidt, W. 1969; see also Chapter 4, pp. 122-126. 9. Boyle 1970a, 1971. 10. Tomlinson, R. F. 1967; see also Appendix. 11. A possible exception is that some real-world data elements may actually have the form of grids, and may be represented by a grid.

### Volume of Related Data

The second variable of spatial data handling systems is the volume of pieces of related data that can be carried at any one location within one set of location identifiers. This variable is the second axis of the record plane (the other axis being that of the location identifier).

The unit of measurement of volume is important, because there is a difference between binary bits of storage, which represent the most efficient form of compaction and hence the true value of data volume that has to be handled by any system, and pieces of data, which are recognizable elements of real-world data. The relationship between the two is straightforward and can be expressed as the total number of classifications or groupings used (or at least the next highest binary number that will hold that number of classifications) for each category or piece of data or data element, that is:

$$\frac{\text{Category } n}{\text{Category } 1} \times \text{Classification Groups} = \text{Magnitude of Bits of Data}$$

With the relationship expressed, pieces can be used as the units to subdivide the axis representing the volume of related data, because the relationship between pieces of data and the content of actual systems is considerably easier to understand. The subdivision of the axis is illustrated in Fig. 6.3.

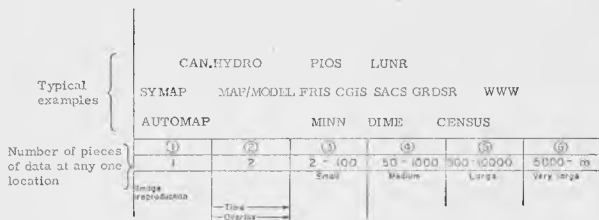


Fig. 6.3. Categories of volume of related data with examples of typical systems.

In the illustration, six categories have been identified. At the lower (left) end of the axis there are two clearly distinguished categories. Farther along the axis the picture is less clear, both as to significant groupings and as to the

exact data content of any one system. In this part of the axis, overlapping ranges have been used to distinguish small, medium, large, and very large volumes of related data. The progression is certainly not linear, and even in this simplified illustration it can be seen to be approximately logarithmic.

The first category contains one piece of data per location. This represents an entity that stands by itself, such as the presence or absence of snow at a particular location or, more commonly, the presence or absence of a road at a particular location. This category includes the commonly recognized image-holding bit-plane found in image reproduction systems such as AUTOMAP, and also the grid cell systems restricted to one value per grid cell such as SYMAP.

The second category comprises records that can contain two pieces of related data per location, and from the viewpoint of system content it is the first category where pieces of data can be related to one another, at one location. It becomes possible to consider time in the data string, and to manipulate overlays. Types of systems within this subdivision include the Canadian Hydrographic System<sup>1</sup> and simple versions of the MAP/MODEL system, although the latter is capable of extension into the third category.

The third category contains data records with a small number of pieces of data at each location, typically 2 to 100. Into this subdivision come such systems as FRIS, NCC,<sup>2</sup> OEM, PIOS, and the CGIS in its current status, although the last of these systems has the capacity to expand into the fourth subdivision.

The fourth category has a medium number of pieces of data associated with each location identifier, typically 50 to 1,000. Included in this category are the data-intensive census types of records such as the GRDSR system, DIME, SACS, and the LUNR type of system.

In the fifth category, large numbers of pieces of data are associated with each location identifier, typically 500 to 10,000. Examples of this subdivision are traditional census-data handling systems, where large numbers of variables are gathered for areas located by name. This category is also characterized by data strings where the attributes are measured at repeated

1. Boyle 1970a, 1971. 2. Symons 1972a.

time intervals, thus producing large volumes of data. An example of this type would be the information attached to one location such as a weather station on a 10-year series of weather maps, each produced every hour. The existing world-wide weather system has data records in this category.

The sixth category has a very large number of pieces of data associated with each location identifier. The range is suggested to extend from 5,000 to infinity. No data system known to the author currently handles data volumes in this category, but if many large files of data were integrated on a spatial basis such volumes might conceivably occur in the future.

The six categories group various types of systems together and show in sequential order of magnitude the impact of the related data volume on the file structures that must be incorporated into the various systems. No attempt has been made at this stage to evaluate the degree of impact of related data volume on system design. It is probably not linear.

#### Manipulation Facilities

The third variable is that of the manipulation facilities incorporated in the system. To establish a hierarchy of manipulation facilities, it is necessary to establish some relative measure of the effort needed to provide them, that is, some measure to identify which manipulations are easy and which are more difficult. For each manipulation, one or more logical operations must be performed upon the data to produce the desired results. A logical operation is a change in data value, a comparison, or movement of the data element. A hierarchy of manipulation facilities can be based on the increasing number of logical operations they contain.

A distinction should be made at this point between a logical operation and a physical operation within the computer system itself. A change in data value, for example, can be accomplished by any of the physical operations of addition, subtraction, multiplication, or division, either singly or in sequence (multiplication and division are in themselves a sequence of the physical operations of addition or negative addition). The result of the sequence of physical operations, however, is a logical operation, the change in data value.

The series of manipulation facilities are laid out along the third axis of the



cube, in Fig. 6.4. Three significant categories of system can be identified: those without overlay capability, those with overlay capability, and those with a significantly developed monitor system or query language, or both.

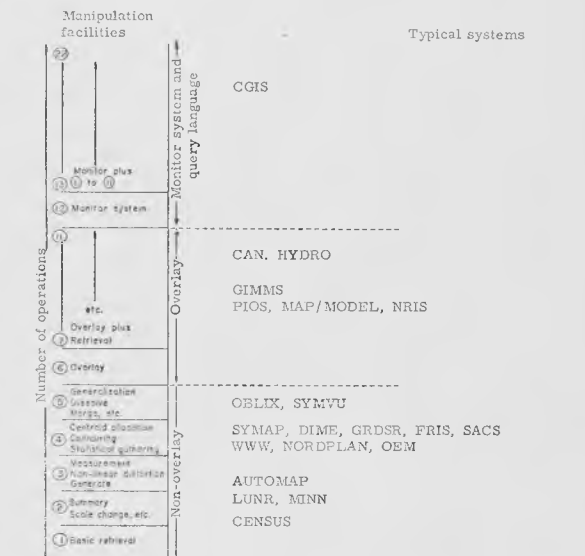


Fig. 6.4. Manipulation facilities and their related systems.

Level 1. Within the first category, the hierarchy of data manipulation can be laid out. The simplest capability (level 1) is basic data retrieval, a single logical "Move" operation. The data content is not manipulated but is merely reproduced on an external storage medium, which could be a piece of paper.

This first example illustrates the difference between logical and physical operations: the logical Move operation may require several physical operations. These can include: Define what is being looked for; Read the input request information; Analyse the content of the request; Move from external

medium to storage to allow comparison internally; Search the tape to find the information (Search itself is made up of Read, Compare, Increment, Move tape, Repeat; when equal; Compare and Move element); Determine what piece of external storage has to be referenced (card punch, printer, etc.); Re-format as required to satisfy the request, (this is usually a complex Move operation, although it can be done as part of a sophisticated Print operation); and Output (print, punch, or display).

This total basic retrieval process may take between six (PL/1) and approximately 200 (Assembler language) instructions, depending on the level of computer system and the language being used. In larger and more advanced computers, particularly those with virtual memory operating systems, the external memory is in many ways an extension of the internal storage and the instructions required are greatly simplified. However, the process followed in basic data retrieval is that of reproducing, on an external medium, material that already exists on an internal medium. This is described as a logical Move, and is one logical operation.

Further levels of capability for manipulation will be described as combinations of such logical operations. In each case the number of logical operations specified may be repeated as often as necessary to complete the manipulation. At the risk of not identifying subtle differences in manipulation facility, and to simplify discussion dividing the capabilities for manipulation into only a few levels, operations that are logically the same or that could logically be done in one step will be grouped as one logical operation.

Level 2. The second level of capabilities is the result of two logical operations, and these can non-interactively provide data summary, elimination of linear distortion, classification change, selective search, scale change, projection change, or measurement of straight-line distance between points. These facilities are described below; it should be noted that the manipulation facilities must remain non-interactive to be classified at level 2.

Data summary is achieved by the combination of the logical operations Move (to acquire data) and Change in Data Value. Elimination of linear distortion, the revision of the location of spatially distributed points in a linear fashion

on one axis, is achieved by a logical Compare operation (against standards) and a logical Change in Data Value. The latter is usually a complex operation of multiplying and dividing, which produces a shift, on one coordinate at a time, from one position to another. Classification change is brought about by a logical Compare (of the existing data values with a table of data values, to find the matching entry for the subsequent substitution) and a logical Move (to effect the substitution). Selective search is carried out by a logical Move to get the data, followed by a logical Compare (against the request). Scale change and Projection change, like Elimination of linear distortion, are a combination of logical Compare and logical Change in Data Value. The measurement of the straight-line distance between two points is achieved by a logical Move to acquire the second point and a logical Calculate (Change in Data Value). A logical Calculate is the same as a Change in Data Value where the change is made between two variables or between a variable and a reference value; that is, it is a parametric change in data value.

The ability to carry out more than one of these manipulations non-interactively establishes a subhierarchy within level 2. However, as long as the manipulations are non-interactive, the system is still considered to be within the level. The same subhierarchical grouping can be applied within other higher levels. Traditional census systems involving summary, LUNR, and the Minnesota System (MINN)<sup>1</sup> can be considered as examples of level 2 systems.

Level 3. The third level of capabilities is a result of three logical operations. These provide for measurement, elimination of non-linear distortion, the ability to generate (circles and polygons), and the ability to carry out statistical gathering.

Measurement can consist of area calculations, determination of distance along an irregular line between points, perimeter measurement, and the count of points within areas. In area calculation, the actual measurement process is assumed to start when the data have been found on the appropriate file and a vertex on the boundary has been identified as an address of the area location. A typical and straightforward area calculation (CGIS<sup>2</sup>) uses the standard algorithm for subtended areas. The first operation is a logical Compare (at the given starting point) with the boundary definition, to determine the next

1. Borchert 1969. ' 2. See Appendix.

point on the boundary. This may be a logical Move, depending on the structure of the data storage; it is so in the case of CGIS, where each vertex would contain the direction and the distance of the next vertex on the segment. The comparison would, however, be implicit on the move to the next vertex. The logical Compare is followed by a logical Calculate (of the area subtended), and another logical Calculate (to create the total).

The distance along an irregular line between two points, or a perimeter, a special case where the line forms a closed loop, is obtained by measuring the length of the straight-line segments that make up the irregular line and summing the results. Both manipulations are made up of a logical Move (to get the position of the end of the straight-line segments); a logical Calculate (of the straight-line length of that segment); and a logical Calculate (to produce the total). As the manipulations become increasingly complex, they may be achieved by several sets of logical operations rather than by just one set. In this discussion, a manipulation will generally be described in a simple form, to indicate the level in the hierarchy of manipulations at which that particular kind of manipulation becomes possible.

A count of points within an area (polygon), another function of measurement, can be achieved in various ways. The simple topological technique is to move from the given point in a uniform direction to an axis, and count the number of intersections made with the boundary of the prescribed area. This algorithm is usually made up of a logical Move (to get the next point); a logical Calculate (of the number of intersections crossed going to the axis); and a logical Calculate (to count the number of points). Different approaches to the count of points within polygons are used in various systems. The digital approach to count the points within the polygon, used in the MAP/MODEL system, and the approach used in the CGIS to locate the area into which a point falls rather than to count the points within an area, both require three logical operations.

The Generate facility is the ability to produce circles, polygons, and points that must be known to the system. The process of generation itself assumes that the position of the item to be generated is specified. Thereafter, the generation of the item proceeds through the logical operation Move (to acquire

the data, including the specified position); a logical Calculate (to determine the distance to the boundary from the item concerned); and a logical Calculate (to determine the movement to the next vertex point). Repetition of this process generates the required figure or point and places it in storage for further use.

Statistical gathering is the process of selectively searching for items according to a given profile, and the application of a specified statistical calculation to those selected. The manipulation can be described as a logical Move (to acquire the data); a logical Compare (against the profile provided in the request); and a logical Calculate (which is a change in value according to the rules provided in the request).

A system in level 3, or any other level, may not automatically include all the manipulation facilities contained in the previous levels, but the manipulations it can carry out are more sophisticated. The AUTOMAP system may be considered to be in level 3.

Level 4. The fourth level of capabilities is the result of four logical operations used in sequence, and these provide the manipulation facilities of centroid allocation (because measurement effectively has to be done first) and automatic contouring. Typical systems within level 4 include DIME, OEM, SACS, FRIS, NORDPLAN, and GRDSR.

Just as there are several ways of calculating area other than the single algorithm mentioned in level 3, there are several ways of calculating centroids. The choice and efficiency of procedures for calculating areas depend on the method used to store the lines as boundaries of areas within the computer. The difference between approaches by which lines are represented as formulas of curves, or as straight-line segments, or as closely spaced points or small grids, can be readily envisaged. The choice of line storage method, and hence the approach to area calculation, is dependent on the character and volume of the original line data, balanced against the need to preserve the accuracy of those data and the cost of computer processing. Centroid determination requires an extra calculation, that of dividing the total moment by the total area; the additional logical Calculate operation involved places it in

level 4. The logical operations involved are: a logical Move or Compare (to acquire the data); a logical Calculate (of the subtended area and moment); a logical Calculate (to sum the totals); and a logical Calculate (to divide the totals one into the other to provide the coordinate).

Automatic contouring is the facility of creating isolines through a set of points with variable values. The simplest form of automatic contouring is the process of starting at a point with a specific value, or a value within a specified range of values, then joining it to the closest point within the same value range. There are more sophisticated approaches to automatic contouring, but this elementary one uses a minimum number of logical steps and is described to indicate the lowest level in the hierarchy of manipulation facilities at which automatic contouring becomes possible.

Assuming that the process starts at a given point, the steps required are: retrieval of the next point to be considered; determination of the distance between the given point and the next point; comparison of the value of the next point with a given set of rules to determine if it is the closest point within the same range of values, to check whether it has been used before, and to ensure that the resulting contour will not cross another contour already in existence (though it can touch); and finally to join the points by attaching a common identifier to the new point.

The logical operations needed are: a logical Move or Compare (to get the next point); a logical Calculate (to determine the distance); a logical Compare (with a set of rules); and a logical Move (to a subset, to assign a common identifier).

The relative simplicity of the set of rules in the logical Compare is the essential difference between this level of manipulation and higher-level manipulations such as the determination of intervisibility between points. In the latter case, the logical Compare would split into a logical Compare/Move process. SYMAP V and WWW, with their capability for automatic contouring, may also be considered to be at level 4.

Level 5. The next level of capabilities is the result of five logical operations used in sequence. This provides facilities for the simplest forms of generalization, merging, dissolving, determination of intervisibility, and simulation. Typical of this level are systems with facilities for hill shading or route

determination.

Generalization is considered to be the process of combining multiple areas, possibly with multivariable classification descriptors, into a single larger area with a single classification. The new classification is based on the proportion of the original separate classifications contained in the new area. Generalizations of line shapes such as coastlines, and particularly the "coastal island" type of problem, are applications of the algorithm; this will be outlined later.

The first part of the process is to measure the size of the areas to be combined, which is a repeated level 3 operation. The next step is to compare the information concerning the location and size of the areas with a set of rules. The comparison will determine if the areas fall into a size category that requires them to be generalized and, if this is so, with which areas and according to what rules the generalization will take place. For example, it may not be permitted to combine certain types of classification, such as land and water, or, if such a combination is allowed, a resulting classification as water may or may not show that the area contains a percentage of land. The Compare operation reflects the rules for generalization that apply in the particular case and forms the basis of the calculation, which follows the prescribed rules.

The first logical operation involved is a logical Move (to get the data). This may instead be a logical Compare of the given starting point with the boundary definition, to determine the next point on the boundary. This operation is followed by a logical Calculate (of the area subtended); a logical Calculate (to create the totals); a logical Compare (with a set of rules); and a logical Calculate (according to the set of rules, to produce the generalization).

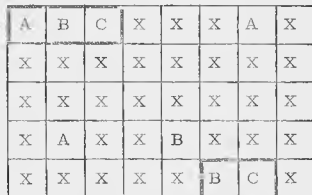
The case of generalization of islands on a coastline follows the same algorithm. The size of the islands is measured as above. The fourth step, the logical comparison with a set of rules, determines whether the islands are smaller than a prescribed minimum size and can be eliminated (replaced by blanks) during the Calculate operation; whether they are large enough to be kept but too small for easy recognition, in which case the subsequent Calculate

enlarges them; or whether they are so large that generalization is not desirable. If the prescribed size limits in the rules were set at a suitably low level, the Calculate operation could make Prince Edward Island the size of Greenland.

Merge is the process of combining contiguous areas that qualify for inclusion in a given new category or set of categories, as defined under manual procedures in Chapter 3 and in the CGIS description in the Appendix. The process is illustrated in Fig. 6.5.

#### Case 1

Merge would occur here



A	B	C	X	X	X	A	X
X	X	X	X	X	X	X	X
X	X	X	X	X	X	X	X
X	A	X	X	B	X	X	X
X	X	X	X	X	B	C	X

Map showing 1-mile-square fields of land classified according to quality.

#### Case 2

Merge would occur here

Given that land types A, B, and C are prime land, and it is required to determine the size and number of pieces of prime land on the map, the analysis would provide the answer that there were five pieces of prime land, one of 3 square miles, one of 2 square miles, and the remainder of 1 square mile. The Merge facility would be used in cases 1 and 2.

Fig. 6.5. Example of the merge process.

The Merge facility would be used for all contiguous occurrences of land types A, B, and C in the figure, such as in Case 1 and Case 2, to combine their pieces for measurement and to assign the resulting pieces to the category "prime land". This process is similar to generalization in its logical path, but is slightly different in application.

The first part of the process is to calculate the size of the areas to be



considered. (Although the calculation is part of the logical process, it may have been done at an earlier stage in some systems, such as CGIS; this shortens the subsequent Merge process and makes it more efficient.) The information describing these areas is compared with a set of rules, to determine if the classifications are of a type that can be merged and if they are contiguous with similar areas. This leads to the calculation in which the sizes of the areas to be merged are summed.

The sequence of logical operations is: Move, Compare, and Calculate operations (to measure the areas); a logical Compare (against a set of rules); and a logical Calculate (to add the totals and create the new combined areas).

Dissolve is the process of removing a common boundary between specified areas and grouping them into a common area. It is simply a single merging of two specified areas. The process of Dissolve and the process of Merge are identical, except that the application of the Dissolve to the actual image includes in its calculation the removal of the identified boundary and the recalculation of the values of pointers on the segments, as directed by the set of rules.

The concept of determining intervisibility between given points is straightforward. The aim is to find out if any intermediate point on a line between two given points has a higher altitude than the plane on which they are located (Fig. 6.6).

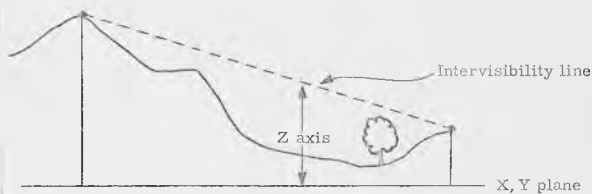


Fig. 6.6. Intervisibility between given points.

The steps in the process are to retrieve the data; to calculate the coordinate on the X, Y plane of the next point (obstruction) to be processed; to calculate the Z coordinate of the intersection; to compare the results with a set of rules to determine if the next point is close enough to the intervisibility line

to be considered, and if so if its Z coordinate value is higher than that of the intersection; and to move an identifier to assign a value to the intervisibility line if the comparison shows that it is equal or higher.

The logical operations are: a logical Move (to get the data); a logical Calculate (to determine the distance away from the line between the two points); a logical Calculate (of the height of the intervisibility line at that point); a logical Compare (against a set of rules); and a logical Move (to identify the line).

In the simplest terms, Simulation is the process of taking statistics and extrapolating trends from them under certain constraints and probabilities. This is a very basic viewpoint of a process that can be extremely complex. Simulation differs from the manipulations described so far, in that the others are essentially sets of predefined operations in a specified sequence, whereas each simulation is a unique interpretation applied to each set of data. To achieve an optimum simulation for a particular purpose, one must create a special set of operations that produces the desired result for the specific data set. Defining the algorithm is part of this process. For the sake of economy, prespecified simulation models are often used on various sets of data, but this usually requires a sacrifice to generality and a compromise in some specific aspects of the problems to be solved.

The basic approach to simulation can be thought of in the following steps. First, the statistics that describe the constraints and the problem have to be gathered together. The second step is the generation of an algorithm (formula) to represent those statistics. The third is the generation or selection of probabilities or probability distributions that will best represent the statistical trends provided in the data gathered in step one. The fourth is to apply the probabilities to the formula, to generate the forecasted result. The logical operations needed to achieve these steps are: a logical Move (to acquire the data); a logical Compare (against a profile provided in the request); a logical Calculate (according to the rules in the request); a logical Calculate (to generate the applicable probabilities); and a logical Calculate (to apply probabilities to the formula, to generate the forecasted results).

Depending on the algorithm that has been developed, the final step is actually the "simulation" step.

Level 6. The second major category of manipulation facilities contains those that incorporate overlay (see Fig. 6.4), and level 6 is arbitrarily occupied by the overlay facility itself. Overlay is the well-understood ability to superimpose two maps (that is, two sets of images and classification systems) and sensibly combine them into one map (that is, into one image and one classification system). The general principles of thematic map overlay will be described below, and although there are several approaches to the task, the one used in CGIS will provide examples where details of a specific algorithm are given.

A preliminary step in the process of overlay is to select the particular parts of the two maps that have to be overlaid, in other words, to define the subsets of the total area that have to be overlaid. Fig. 6.7 shows Map B partially overlaid on Map A. The decision must be made whether to include in the output all portions that are common to A and B ( $A \cap B$ ), or to include all areas of both A and B ( $A \cup B$ ). In the latter case, an arbitrary null classification could be used for the portion of the resulting overlay that does not overlap; in the CGIS, character blanks are assigned. A further choice is to overlay only certain portions of the intersection of A and B.

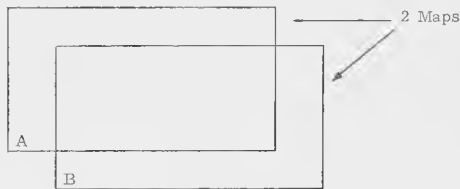


Fig. 6.7. Overlay condition where Map B is partially overlaid on Map A.

The first step in the actual overlay procedure is to lay out the image of the first map in memory. The image of the second coverage is then laid over the first image in memory. During this process the images are compared, point by point, for intersections of boundaries and for each intersection point found, a value (the X,Y coordinate of the intersection) is put into a table. Next, the table of intersections is sorted into a convenient sequence for

subsequent comparisons; in the CGIS approach the table is ordered by X, Y values. The number of boundary intersections so stored is theoretically infinite, and is limited only by the capacity of the memory system and the fact that a meaningless quantity of intersections may be created when numerous coverages are so overlaid. The CGIS design has arbitrarily set eight medium- or low-density map coverages as its upper limit. The overlay of two map coverages with extremely dense boundary sets can cause problems with memory capacity at this point in the process.

The next operation is to re-examine each of the two coverages being overlaid, and to break up their boundaries into subsegments whose ends are the same as the newly determined intersection points. The subsegments are then reconnected to create the new areas and so complete the overlay of the images. This last step must be done sensibly with regard to the classifications attached to the previous boundaries.

Systems do not necessarily keep their image and descriptor data sets in different files, as in CGIS, and the combining of areas and classification systems sometimes proceeds simultaneously. Similarly, there are different approaches to the process of combining the classifications from two or more maps to produce the classification for the new map. Generalization may be employed. In combining 12 weather maps, each showing the average mean temperature for a month of the year, one may wish to produce an end result showing areas with a specific profile of temperature, for example, an area that enjoyed 30% of the year between 30 and 40°F, 40% of the year between 40 and 70°F, and 30% of the year between 70 and 90°F. This type of change in classification could be handled by the image overlay process described above, but the generalization of the descriptor data would have to follow a complex set of rules that would depend to a great extent on the format and nature of the data. The CGIS approach is to use simple concatenation of the descriptor data sets, so that the descriptor of each new area formed by the overlay is obtained by simply adding together the classifications of the two original areas. The process follows a simple set of rules and is not data dependent. The total classification of one area is combined with the total classification of the overlying area. There are, of course, variations in

the approach to overlay, depending on the nature of the computer being used, the size of the records being handled, and the results required by the user.

The logical operations involved in preliminary selection of the areas to be overlaid are: a logical Move (to get the data); a logical Compare (against a set of rules); and a logical Move (to a subset). In the basic type of overlay process, the logical operations begin with a logical Move (from auxiliary storage to memory) to lay out the image of the first coverage; the second image is laid out and the subsequent comparison and formation of the table of intersections are done by a repeated logical Move and logical Compare, followed by a logical Move (of the identified intersections to the table). The table is sorted into sequence by a cycle consisting of a logical Compare followed by a logical Move. The existing boundaries are broken down into subsegments by a logical Move (within the image data), a logical Compare (with the already sequenced intersection table), and a logical Move (of the data subsegments to storage when the next intersection is reached). The subsegments are reconnected into areas by a logical Compare (to make sure that it is another segment bounding the same area) and a logical Move to redirect the pointers.

Levels 7 to 11. The sequence outlined above was taken mainly from the process of image overlay used in the CGIS, and 11 logical operations were identified in it. An overlay process based on images alone could take somewhat fewer logical operations, but at the expense of processing time, or if based on a less complex data structure. One with more sophisticated handling of complex image conditions or of descriptor data could have more operations. As the overlay manipulation is considered to be significantly more complex than the manipulations described in levels 1 to 5, it has been arbitrarily assigned to level 6 and forms the basis of the second category. The levels described in the first category can be repeated as levels in the second category (with the addition of an arbitrary six operations on the front of each manipulation), providing levels 7 through 11.

Level 8 is thus a system with overlay capabilities that can also carry out manipulations up to level 2, either before or after the overlay manipulation.

Level 9 incorporates overlay capability within the facilities of level 3, and is typically represented by the PIOS, MAP/MODEL, NRIS, and GIMMS systems.

Levels 12 to 24. The third major category incorporates all systems with a monitor system or query language, or both, that have been developed or implemented to a significant degree. This facility may exist with or without the overlay capability and, using a base of 12, provides levels 13 to 24. Its position on the axis (see Fig. 6.4) thus need not be a linear progression from the overlay; it is placed there merely for convenience, and perhaps to indicate the significantly high level of capability for manipulation. The sequence of levels could be continued through this third category to indicate the relative sophistication of capability for manipulation. Although the position of systems with monitoring capabilities on the axis is arbitrary, in practice they have not been considered necessary for systems with relatively simple manipulation facilities. Only the higher-order systems incorporating overlay or interactive processes, or both, employ this increased level of sophistication. The progression of manipulation facilities along the axis thus represents a relatively clear progression of actual capabilities for manipulation in systems in 1974.

A monitor system and query language facility is not an operation in itself; rather, it is an overall controlling function of other operations. A monitor system is the executive director of normal operations within the computer and its purpose is to free the operator from manual control tasks; it makes it as easy as possible for the user to have his request answered. A query language is a set of code words used to specify the manipulation and retrieval functions required. One of the basic functions of the monitor system is automatically to combine manipulation or retrieval commands from the query language with the descriptive information in the appropriate files, and with computer programs resident in the system libraries. It thus generates a complete request that the computer can use to enter the data bank, produce the desired manipulation, and present its results in tabular or plotted form. A full technical description of a typical monitor system and query language is given for the CCHS in the Appendix.

The logical operations involved in the basic theoretical approach used by the

monitor system and query language are as follows: a logical Move (to read the attention signal from the input device); a logical Move (to send acknowledgement to the device); a logical Move (to read the initialization record from input or output of the last completed task); a logical Compare (to check the validity of the user, or against the table of active tasks); a logical Move (of a message to the terminal telling it to enter the request, if the request is from a terminal, or to remove the task identifier from the table of active tasks); a logical Move (to read the first line of the request, or, if this is the last step of the process, to return to waiting status); a logical Compare (of key word with list, to edit the key word for validity); a logical Move (to read file definitions on a specially designed system file associated with the monitor to determine the files associated with the program); a logical Move (to call in the optional parameters provided in the request to modify the file definition, for example, the names of the coverages to be accessed); a logical Move (to load the program from the library to the computer); a logical Move (to enter the task identifier into the table of active tasks); a logical Move (to hand over control of processing to the program and so initiate the task, and then return to waiting status); and a logical Move, Compare, Move, Calculate, and Move sequence (see query language). The above logical steps are then repeated as necessary.

The only system known to have a full monitor system in conjunction with overlay and all lower-level capabilities for manipulation is the CGIS, which would have an approximate level of 22. The USGS system,<sup>1</sup> now in the experimental phase, and developments of the U. S. Engineers Waterways system<sup>2</sup> anticipate attaining this level of manipulation facility.

The above discussion of manipulation facilities has been more extensive than those for the axes of location identifier and volume of related data, mainly because the techniques involved have not been classified before and it was necessary to establish some basis for comparison, however broad and simplistic. The categories and subcategories identified (see Fig. 6.4) are not linearly distributed along the axis. There are obvious discontinuities of scale among the three main categories, and the distribution is not arithmetic but tends to be logarithmic in terms of the increase in complexity of

1. Edson 1973. 2. LaGarde 1973.

manipulation along the axis. A more detailed examination of the processes would probably reveal that the progression was multi-dimensional. For the purposes of this discussion, however, the axis provides a simple empirical summary of the relative levels of manipulation capability in systems.

#### SYNTHESIS

The following series of diagrams illustrates the location of the various systems in relation to the axes described above.

Nineteen systems have been located on the following diagrams. Not all the systems mentioned in the previous chapters have been plotted, as this would have incurred many closely overlapping names and unnecessary duplication. (The SYMAP system, for example, is representative of approximately ten similar systems that would each occupy approximately the same space on the diagram.) However, each system previously mentioned in the text as an example of a particular category of storage format is plotted. Where space permitted, and differences warranted, other systems have been added. The Minnesota (MINN) system, which uses 40-acre grid units, has been plotted as well as the LUNR system, which uses 1-km<sup>2</sup> units. The "Census" system represents traditional, computer-aided, census handling techniques used prior to the advent of the GRDSR and DIME type of systems, and which still account for the greatest amount of census data handling. The world-wide weather (WWW)<sup>1</sup> system, which is essentially a point data system, has been shown as it represents the largest numerically coded, location-specific data set currently being handled. Manual techniques are not identified as they are not traditionally recognized as systems. Their contribution to overall data handling capability will, however, become apparent.

The first diagram (Fig. 6.8) locates the 19 systems in the record plane, with respect to the variables of location identifier and volume of related data. This perhaps is the most interesting of the three views of the cube shown, as it gives an overall picture of the general availability of systems (regardless of manipulation capability) to handle various types of data.

1. Personal communication, McKay 1973.



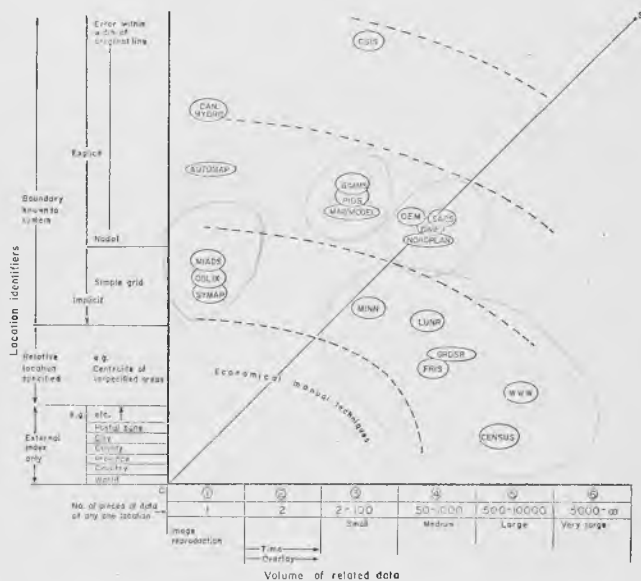


Fig. 6.8. Location identifiers related to volume of related data.

The area of the diagram closest to the origin of both axes contains no computer-aided systems. In this area of simple location identifiers and relatively low volumes of related data, current manual techniques of storing and analysing data are economical and satisfactory to the user.

Immediately beyond this area is a proliferation of current working systems, located according to their need to handle images in related data sets. The various arbitrary regular-area grid systems such as SYMAP, MIAD<sup>1</sup>, MINN, and LUNR are represented here, as are the point data systems of GRDSR and FRIS.

These systems are reasonably well established in current practice, but it is

1. Amidon 1964, 1966.

interesting to note that as increasing volumes of data are handled, a retreat is being made into the simpler forms of location identifier. The progression of grid sizes in relation to data element size increases (SYMAP-type, to MINN-type, to LUNR-type) as volume of data increases. Point data systems such as GRDSR, FRIS, and WWW, utilizing location identifiers that provide only the relative location of their data sets, are used as data volume increases.<sup>1</sup> With the traditional techniques for handling total sets of census data it has been found essential to limit the location identifiers to nominal codes.

Beyond the well-established systems is a band of others that are generally operational, but still somewhat experimental. The AUTOMAP system comes into this category by reason of its attempt to handle simple images with a relatively high order of location identifier. The DIME, OEM, NORDPLAN, and SACS type of network systems fall into this category because, although they utilize a lower order of location identifier, they attempt to handle medium volumes of data. The coarse-polygon systems of the MAP/MODEL, PIOS, and GIMMS type fall into this category because they combine the first attempts at using polygon location identifiers with small amounts of related data.

The Canadian Hydrographic System and CGIS represent systems under development, which attempt to handle data with a high degree of location specificity. It should be noted that this is the reason why they are relatively high-order systems. Neither handles large volumes of related data at present. The Canadian Hydrographic System is primarily an image reproduction system. CGIS still handles relatively small volumes of data at any one location, although it may handle considerably more as its development continues.

The next diagram (Fig. 6, 9) shows the relationship between the manipulation facilities of the various systems and the type of location identifiers used.

1. Although GRDSR and FRIS are designed to represent areal data by a point data surrogate, it could be suggested that point data collected at weather stations, as in WWW, are precisely located in a point data system. This suggestion would, however, be misleading, as the weather station data are by definition taken as a sample of the areal atmospheric conditions and indeed are the basis of subsequent areal analysis. The constraint imposed by the data-gathering process does not alter the fact that areal data are being surrogated by point data for the necessary purpose of data reduction.

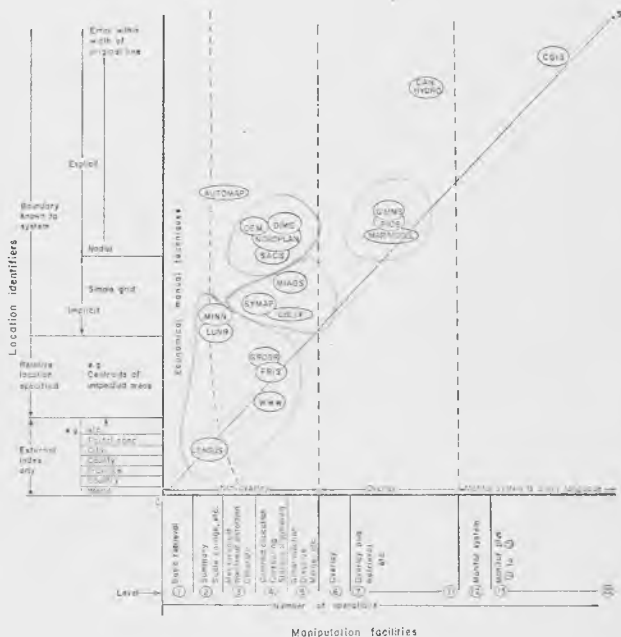


Fig. 6.9. System manipulation facilities related to location identifiers.

If one considers the human mind to be the equivalent of a monitor system and query language combined (and from the conclusions reached in Chapter 3, it performs these functions with considerable efficiency), then, disregarding volume of data, manual processes may be said to be applicable to any type of manipulative activity, in concert with any type of location identifier encompassed in the above diagram.

Because of the limitations of human motor capacity and human channel capacity, however, only the simplest of manipulative functions can be

handled economically by manual processes. For even limited volumes of data, the evidence presented in Chapter 3 indicates that manual techniques can only be considered economical in approximately the first two levels of manipulation facility identified in the diagram. (This, however, has not prohibited their application to very small volumes of data for higher-order manipulation procedures.) The development of computer-aided systems has thus seen its greatest proliferation in levels 3, 4, and 5 of manipulation procedure, as may be expected. These essentially represent the immediate extension of manual processing capability. From the distribution of the systems on the diagram, it can be seen that traditional census handling techniques, and the larger grid systems LUNR and MINN, overlap into the area of economical manual techniques. They compete directly with processes that are still manually efficient (assuming limited data sets). AUTOMAP can carry out search, scale change, and projection change on explicitly defined boundaries, but these activities appear to be at the limits of manual efficiency even though they are carried out manually with some regularity.

An identifiable group of systems falls well within levels 3, 4, and 5 of manipulation facility. These processes are demonstrably expensive to carry out manually, and it is to be expected that computer-aided techniques would be first developed for such facilities. Ten of the 19 systems shown fall into this limited area. It is interesting to note, however, that the range of location identifiers employed by such systems is still quite limited. They use point data sets, small grids, or simple network representations as their basic data formats. Little capability exists to perform the higher manipulation functions of the first category on the more explicitly defined spatial data.

Systems incorporating an overlay capability necessarily occur in the upper half of the location identifier axis because, unless boundaries are explicitly identified, the union of data sets can be accomplished by summary. The relatively small number of systems incorporating this capability (five out of 19) suggests that either little requirement exists for the results of overlay, or the process requires techniques that are mainly experimental at present. The truth probably lies in a combination of these factors, although the



increasing volumes of related data on even the simplest manual manipulation processes. In Fig. 6.8 (systems in the record plane), it was suggested that traditional computer-aided census techniques, and the larger grid-based systems (LUNR and MINN), were not economical to perform manually.

In Fig. 6.9 (manipulation facilities related to location identifiers), they appeared to be competing with manual processes, assuming limited amounts of data. Fig. 6.10 shows that the utility of these simple approaches to spatial data manipulation rests on their economy in handling medium to large amounts of data rather than on their resolution of location identifier.

Toward the left of the diagram, in the first category of volume of related data, the image reproduction systems can be identified. AUTOMAP appears to be a low-order system in these terms, although it handles a high-order set of explicit location identifiers. The highest order of the image manipulation systems shown is undoubtedly the Canadian Hydrographic System, combining as it does a high capability for manipulation with high-resolution location identifiers. (Its purpose is to manipulate precise hydrographic charts prepared for navigational purposes.) The middle range of the image reproduction systems is occupied by the Type 2 arbitrary regular-area<sup>1</sup> family (SYMAP, MIADS, etc.) that accepts a low to medium level of location identifier (small grids) but offers a relatively wide range of capability for manipulation in the non-overlay category.

The next group is FRIS, SACS, DIME, OEM, GRDSR, NORDPLAN, and WWW. These systems are characterized by a fairly extensive capability for manipulation in the non-overlay category, and by a varied but medium level of location identifier (see Fig. 6.9); however, all in some way attempt to handle considerable quantities of data related to their location identifiers. Regardless of location identifier, the emphasis is on related data handling. It is significant that, with the exception of WWW (itself a system handling large amounts of data), all systems in this group have been developed in agencies directly related to census and administrative data handling activities. This, then, is the group that has attempted to handle large quantities of data and has compromised on location identifier and manipulation capability (at the moment).

1. As defined in Chapter 4, p.98.

In the overlay category is the well-established group (in all three diagrams) of GIMMS, PIOS, and MAP/MODEL. Here the emphasis has been on the development of capability for manipulation, and in every case the file structure is designed for this rather than for carrying large amounts of data. It is relevant that each system has been developed by planners or geographers concerned with spatial analysis, rather than the manipulation of large data sets.

The CGIS system has a substantial capability for manipulation but at present it does not carry large volumes of related data at each location. The file structure of the system (the separate descriptor data files) is in place to handle substantial quantities of related data in an efficient manner.

In summary, the first diagram of the three (Fig. 6.8) generally illustrates the current status of development of systems and their ability to handle various types of data. The two diagrams thereafter (Figs. 6.9 and 6.10) clarify the first and permit certain group characteristics to become apparent. The image-handling systems, with their varying levels of location identifier and capability for manipulation, can be recognized. This group is probably representative of systems of "automatic cartography" now in embryonic stages of development in various traditional cartographic institutions. The group utilizing low-level location identifiers (CENSUS, LUNR, MINN) can be identified as depending on data volume for their economic application. The larger group of working systems, with non-overlay manipulation facility and some variety of medium-level location identifier, are shown to be those attempting to tackle the problem of handling data volume with greater facility than is offered either by manual techniques or by the systems using low-level location identifiers. They directly reflect the immediate needs of the agencies developing them.

The coarse-polygon group of systems (MAP/MODEL, PIOS, GIMMS, etc.) can be identified, and essentially represents an attempt to achieve the capability for spatial manipulation on limited data sets and to facilitate the graphic tasks that are prohibitively expensive by manual methods.

CGIS is a research system designed to have a relatively high capability for manipulation coupled with the potential to handle medium to large data sets.

It must be emphasized that the diagrams themselves are not, and cannot be considered, a precise tool for relating system capabilities; they merely provide a conceptual framework for discussion and for these general observations.

Considering the three variables as the axes of a cube, as illustrated in Fig. 6.11, the diagonal OS essentially represents the potential capacity of a file structure to handle information in any location-specific information system.

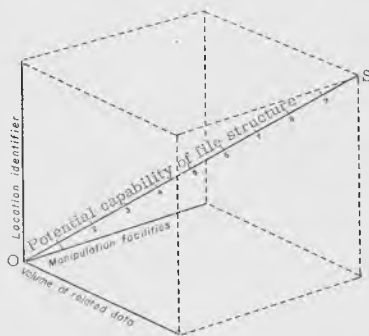


Fig. 6.11. Diagonal OS as a vector of potential capability of system file structures.

If the diagonal axis is divided arbitrarily between O and S into ten equal parts, reference to the three previous diagrams (Figs. 6.8, 6.9, and 6.10) shows that the portion of the cube passed through by the lowest three units encompasses variables that together may comprise the field where current manual methods of storing and manipulating data are economic and satisfactory to the user.

The portion of the cube between levels 3 and 4 on the diagonal contains the proliferation of current working systems according to their need to handle images or related data sets. These probably indicate the level of file structure and design, coupled with machine capability, that has currently



found economic acceptance. Further systems, if developed with combinations of variables that fall into this category, should be implemented without difficulty and be efficient to use.

As one proceeds up the diagonal, the file structures necessarily become more sophisticated; in fact, as the orthogonal axes themselves exhibit logarithmic progression, the diagonal axis is probably logarithmic. The type and capacity of the data processing machine required, as well as the degree of skill needed to use it efficiently, similarly progress. Typical of the systems at levels 5 to 7 are the image-handling systems of AUTOMAP, the Canadian Hydrographic System, and also the PIOS, MAP/MODEL, GIMMS group.

It should be noted that the diagonal represents the potential capability of the structures at these levels. In some instances the file structure is not used at its full potential. The basic characteristics may be employed, but the substructures within the general file organization have not been developed. The structure could be said to be like a skeleton that is capable of supporting a body but does not have the flesh and ancillary systems in place, or does not need them at the moment.

Perhaps one level higher is the type of system represented by CGIS, because such systems attempt to carry more related data than do those mainly concerned with the manipulation of images. At this level, efficiency of operation is closely related to file structure and the optimum use of computing facilities. Here again, however, the potential capability of the file structure is currently under-utilized. Within the general framework of the existing CGIS file structure, a wide variety of substructures could be developed in response to user demands, but they have not yet been established.

Remembering that the diagonal probably progresses logarithmically in terms of increasing complexity, it can be noted that no file structures have been developed in the higher parts of the range. Such systems demand sophisticated design, and probably the use of computer facilities not currently available. The area, however, encompasses systems that must manipulate large volumes of data with high-resolution location identifiers. The data sets to be handled in this manner are not yet in machine-readable form. Also, the implementation of such capability must await the development of

considerably improved processes of graphic to non-graphic data conversion, or the development of advanced computing capability, or both. Even more fundamentally, it awaits the need of users to handle such data sets.

Although this approach to geographical information systems creates some understanding of the relationship between characteristics of the various systems, its primary purpose is not to establish a hierarchy of system sophistication. Rather, it allows the systems to be examined in terms of available environmental data and the effort required to handle them. Just as the variables of location identifier, volume of related data, and manipulation facility may be used to describe the systems, so may they serve to indicate the requirements for handling data to provide a better description and understanding of the environment.

By extension, the same type of cube or matrix used as a framework for system description could be used to examine the data handling requirements of the user, and the relationships between system capability and environmental data handling needs. Further, it may be possible to relate the capabilities of geographical information systems to the capabilities of data gathering processes. In similar terms the stated requirements of the user may be related to the various types of data analysis, and to the decisions that depend on such analyses. These linkages, and hence the overall utility of spatial data handling systems, will be examined in the following chapter.

## CHAPTER 7 - GEOGRAPHIC INFORMATION SYSTEMS, SPATIAL DATA ANALYSIS, AND DECISION MAKING IN GOVERNMENT

In general, the five previous chapters of this thesis have had a technical emphasis. Relatively little discussion has been provided about the theory on which spatial data analysis is implicitly or explicitly based, about the institutions that have fostered the development of various data handling systems, or about the general socio-economic context in which geographic information systems have evolved.

Processes of data gathering, characteristics of data, and techniques of manipulation are inextricably linked with methods of data analysis, institutional structures and activities, and the need to make decisions about real-world problems, even if their interdependence is not clear or completely understood. This chapter will briefly examine some aspects of these interrelationships in order to illustrate the overall value of geographic information systems as they are known today, and to show that they have grown out of a combination of pragmatic need and intellectual curiosity as much as from any narrowly defined progression of technical evolution.

The conceptual extent of the subject, however, is enormous. It would, for example, be quite proper to consider the large-scale social and political background<sup>1</sup> of institutions, in order to understand their priorities and actions in information processing. It would be valid to trace the impact of data on decisions in the various institutions within different social and political structures. It would also be valuable to assess the validity of the theories consciously or subconsciously applied in the current gathering of data (the theories underlying the measurement, sampling, classification, and recording of data).<sup>2</sup> These are important and necessary topics for study and must be examined if the value of information systems is to be fully understood. Any discussion of these broader issues in the concluding chapters of the thesis, however, would necessarily be superficial and unsatisfactory.

The thesis has been concerned with the handling of spatial data gathered by government agencies; therefore, the discussion in this chapter will focus on

1. See, for example, Tomlinson, R. F. 1971. 2. See, for example, Coombs 1964, and the discussion in Harvey 1969, Chapter 17.

the contribution of systems for handling spatial data to the analysis of spatial data and decision making in government.

The concept of the relationships between the three elements is expressed in simplest terms in Fig. 7.1.

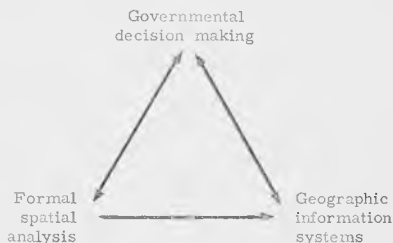


Fig. 7.1. Relationships between data, analysis, and decision.

The circular nature of the diagram immediately emphasizes the difficulty of breaking into the problem at any one point. The availability of data constrains analysis, yet analysis (or the theory behind it) should be a guide to what data are made available. In terms of decision making, the analysis required depends on the decisions that must be made, and the decisions in turn can be affected by the analytical capability. Decisions can be made directly from data (assuming the existence of some mental model), and the concerns of the community and the nature of its institutions can influence the data actually gathered. The translation of this simple conceptual diagram into a model relating techniques of data handling to spatial theory and the practice of institutional decision making is not straightforward. There are difficulties in using a rigorous language to abstract the relationships, in stating reasonable concepts that are precise and unambiguous, and in developing precise statements of the relationships that bind the concepts together. The diagram can thus be used only as a framework for discussion in a preliminary examination of the relationships involved.

The characteristics of geographic information systems and the capability of their techniques to handle spatial data have been examined in the previous chapters. The remainder of this chapter will consider the characteristics,

first, of formal spatial analysis, and then of governmental decision making, in terms of their needs to handle spatial data. The various categories of existing geographic information systems identified in Chapter 6 will be examined in this context, the institutional basis of their development will be discussed, and the usefulness of the different types of system will become apparent.

Chapter 8 will consider the future developments of geographic information systems and examine their continuing role within the discipline of geography and decision making in government.

#### FORMAL SPATIAL ANALYSIS RELATED TO DATA USE

Formal spatial analysis, the rigorous explanation of the distribution and interaction of the attributes of places on the earth's surface, is central to geographical methodology. The contribution of the tools of geographic information systems to formal spatial analysis is thus, to some extent, a contribution made to geographical methodology.

After earlier development in several of the physical and biological sciences, the tools available for rigorous explanation in the social sciences have rapidly developed in recent years. In an essay on the interaction of tools and problems in economics, Koopmans<sup>1</sup> noted that "The present phase of economics is indeed one of turbulence and transition in the domain of tools, more so than in the domain of problems and suggested solutions. Perhaps the most conspicuous development is the increasing use of a growing number of mathematical concepts, theories, and theorems in economic reasoning. But several other important developments demand attention. There is the revolutionary increase in the capabilities of computing and data-processing equipment; the increased application of methods of statistical inference, including methods specifically designed for economic problems; the opportunities afforded by systematic application of sample survey methods of observation."

The words of Koopmans in 1957 will seem familiar to the ears of geographers, who have witnessed a similar radical transformation in the spirit and purpose of geographical methodology in the past 15 years. They serve, however, to

1. Koopmans 1957.

place the contribution of computer-aided systems of data handling in context with that of the other recently arrived "tools".

Perhaps a clearer distinction should be made at this point between the emergence of the capabilities of computers to perform calculations and their use for geographic information systems. At the beginning of Chapter 4 it was noted that before 1960, computers were essentially high-speed calculating machines, and not until the early mid-1960s did developments in computer technology allow rapid growth in their use for information handling and the non-graphic storage of spatial data. Yet in 1963, in a cogent review of the impact of the newer tools on geographical methodology, Burton<sup>1</sup> stated that, while "the consequences of the revolution have yet to be worked out and are likely to involve the 'mathematization' of much of our discipline with the attendant emphasis on the construction and testing of theoretical models," and "although the future changes will far outrun the initial expectations of the revolutionaries, the (quantitative) revolution is now over." Certainly, by 1962-63, the tools of mathematics, the techniques of statistical inference and sampling, and the calculating capacity of computers had begun to make a significant impact on geographical methodology. However, at that time not one of the numerous computer-aided techniques for handling spatial data available today and described in this thesis had been developed. Coppock<sup>2</sup> in 1962 clearly saw the limitations of the computers of the time and noted that the "book-keeping" type of computer was probably faster for large-scale data manipulation than the "scientific" computers then available. He commented that "with a machine of larger storage capacity, it might have been possible to print out results..." sorted by geographic coordinates to make map drawing easier. Such developments were anticipated by Tomlinson,<sup>3</sup> Kao,<sup>4</sup> and others, but even the simpler forms of systems with grid storage formats (such as SYMAP<sup>5</sup> and MIADS<sup>6</sup>) did not appear until 1964-66. Tobler's<sup>7</sup> early contribution to experimental automated preparation of thematic maps was published in 1964. The development of a wider range of systems for handling spatial data has been a product of the past decade (1964-74), and even now they must be considered to be in an embryonic stage of development.

1. Burton 1963. 2. Coppock 1962. 3. Tomlinson, R. F. 1962. 4. Kao 1963. 5. Schmidt, A. 1966. 6. Amidon 1964. 7. Tobler 1964a.

Not only were geographic information systems not in the forefront of the quantitative revolution in geography, but also they have, in the main, been developed by agencies more concerned with the handling of data than with the growth of geographical theory. It is not within the scope of this thesis to repeat or enlarge on the epistemological problems of geography or to trace the overall impact of quantitative methods on the development of geographical theory. It is, however, necessary to mention briefly the topic of geographical methodology to illustrate the potential contribution of geographic information systems to formal spatial analysis.

At the outset, it is probably reasonable to state that the ability to handle spatial data will probably have more immediate impact on geographical methodology than on geographical theory. Within a geography defined as "the description and explanation of the spatial differentiation and interaction between attributes of the earth's surface,"<sup>1</sup> there are a number of "strands" of theory, which are very difficult to discuss in the absence of any as yet clearly identified body of geographical theory. Haggett and Chorley<sup>2</sup> have approached the problem by grouping a wide variety of ideas and approaches within a framework of movements, networks, nodes, hierarchies, and surfaces. There are problems of distinctly spatial data associated with each of these "stages", and some of them have been attacked, consciously or not, by the geographic data handling systems described. Some examples at various stages are systems based on point data, on line segments, on irregular polygons, and on other surface-descriptive formats, and systems with the ability to combine attributes within or between data formats and to derive composite data such as the interaction between two places, their hierarchical status, or their involvement in a particular activity. There are, however, difficulties in this simplistic relationship between geographical theory and capabilities for handling spatial data.

As in any discipline, there are many approaches to theory, ranging from the most abstract, deductive, and idealized mental constructs of theoretical physical science or economics, to the most inductive, data-oriented formulation of working hypotheses, for the purposes of experimental science or social studies. The former type of approach often deals with concepts

1. Surface in this sense implying "shell", rather than the mathematical definition of "surface". 2. Haggett 1963, Chorley and Haggett 1967, Haggett and Chorley 1969.

that are currently difficult to measure (such as a "quantum" in physics or "utility" in economics), or that can be mathematically precise in their meaning while being difficult to envisage in the real world (for example, exponential functions, Lagrangian multipliers, or entropy). The inductive approach, on the other hand, depends on formulating generalizations (theoretical statements) in ways that can be tested through models of various kinds. The inductive approach, in fact, depends on the availability of specific data.

To complicate this situation further, both deductively and inductively derived theories can be tested by reference to models that use various types of surrogate data; for example, utility may be measured by the amount of cash people are prepared to pay for a good; interaction is often "measured" through use of specific sets of data on migration, telephone calls, or commuting flows. The validity of such operational tests of theoretical statements obviously depends on how well the data conform with the assumptions of the model derived from the theory. Thus the difficulty in formulating any direct measure of the impact of existing or improved geographic information systems on the creation of geographical theory can be readily appreciated.

Harvey<sup>1</sup> attempts a brief classification of the various approaches to theoretical structures: Type 1, deductively complete theory; Type 2, systematic presuppositions; Type 3, quasi-deductive theories (which includes inductive systematization as one of its subcategories); and Type 4, nonformal theories. Considering specifically geographic theory, Harvey<sup>2</sup> argues that "Given the nature of geographical concepts, the development of formal theory (Type 1) in geography appears to be a restricted possibility," and, "In practice, most of our traditional (pre-quantitative) theorising has been within Type 4 theoretical structures..." He continues<sup>3</sup> by suggesting that "in the present situation, it may well be more advisable to develop clear heuristic treatments of geographical problems than to seek for full formal expressions of theory. For this reason, model-building techniques are much more significant to geographic research and are likely to remain so in the near future."

Following this line of thought, it is reasonable to suggest that the main contribution to be made by geographic information systems to geographical theory and formal spatial analysis is the provision of data for models or partial

1. Harvey 1969, Ch. 7, pp. 97-99. 2. Harvey 1969, Ch. 9, p. 130.

3. Harvey 1969, Ch. 9, p. 134.



models related to inductive theoretical structures. The question that remains is whether models of use in geographical studies require spatially disaggregated data, or large volumes of data, or whether they require sophisticated techniques of manipulation to derive adequate surrogate data from basic data sets; in other words, whether geographic information systems are actually needed. Certainly there has been little communication between those developing models and those developing geographic information systems. Frequently one hears of models being based on existing data, or simply using special surveys to gather data otherwise unavailable.

The overall spectrum of data needs for geographical models has not been examined extensively. Apart from statistical abstracts and reports on data sources, some publications by data gathering institutions have suggested uses for the data they collect. An example of these is a series<sup>1</sup> of publications by the U. S. Bureau of Census on data use for various types of administrative purpose. Similarly, the administrative needs of planning agencies for data have been examined; this work is typified by the papers of Cheshire<sup>2</sup> in Great Britain, Holleb<sup>3</sup> in the U. S. A., Salomonsson<sup>4</sup> in Sweden, and by a major study<sup>5</sup> in Great Britain that examined the needs of local agencies in terms of the decisions they make. The last study did not, however, relate the data to the still "largely experimental systems approach to planning and modelling techniques." At the other end of the scale are the numerous descriptions of individual spatial models that mention the data problems related to their specific task. Some authors have taken a wider point of view. Coppock and Johnson<sup>6</sup> in 1962 made an early examination of data sources in terms of the burgeoning demands of quantitative geographical methods. Recent work is typified by the reports of Cater,<sup>7</sup> Batty,<sup>8</sup> and Willis<sup>9</sup> in the United Kingdom. Of this later work, that of Batty probably approaches the problem most directly.

Batty considers the spatial theory underlying urban planning and generally classifies the various concepts according to their use of data. He recognizes eight categories, based on whether the data used are aggregated or disaggregated with respect to space, attribute, or time. These categories can be grouped and ranked according to the level of resolution of the data

1. U. S. Bureau of Census 1970a. 2. Cheshire 1970. 3. Holleb 1969.  
4. Salomonsson 1969. 5. Ministry of Housing and Local Government, U. K. 1970.  
6. Coppock and Johnson 1962. 7. Cater 1970. 8. Batty 1970. 9. Willis 1971.

employed. For convenience of classification, low levels of data resolution will be referred to as "relatively aggregated" data and high levels of resolution as "relatively disaggregated" data.

The concept of aggregation or disaggregation of the spatial aspect of data generally relates to scale. Relatively aggregated data in the spatial sense would be used in subregional models such as interurban models, whereas relatively disaggregated spatial data would be employed in intra-urban models at local scales.

In terms of time data, models that are static with respect to time are considered to employ aggregated time data; conversely, those that are dynamic with respect to time are said to be using disaggregated time data.

Similarly, the relative aggregation or disaggregation of attribute data refers to the level of detail at which they are classified. In models using population or employment data, for example, aggregation would apply to those that treat the data as a whole whereas disaggregation would apply to models that subdivide them in some way.

The ranking of models is given below. (Batty uses a different format for the arrangement of his eight categories and a somewhat different assignment of concepts to categories, but the general approach follows his work.)

Level 1: All data aggregated in space, attribute, and time

Examples: Rank size rule, gravity theory (principle of least effort).

Level 2: One aspect of data disaggregated; other aspects aggregated

a) Disaggregated spatial data

Examples: Theories of linear form, concentric form, and microscale design, central place theory, Garin-Lowry models of change of land use.

b) Disaggregated attribute data

Examples: Isard's general theory of spatial economics (minimum cost equilibrium), economic base theory, input-output analysis, social theory of human interaction (e.g., Webber), Forrester's model of urban dynamics, Meier's model of human communication.

c) Disaggregated time data

Examples: Migration models (e.g., Hägerstrand).

Level 3: Two aspects of data disaggregated, third aggregated

a) Disaggregated spatial data, disaggregated attribute data

Examples: Socio-ecological theories (land use - land rent equilibrium) such as the concentric zone theory, multiple nuclei and sectoral theory, economic rent, theory of agricultural location and urban location, Wingo's model of urban land use, Muth's model of residential location, Kain's model of residential location, Herbert-Stevens' model of residential location, Chapin-Weiss' model of residential location, etc.

b) Disaggregated spatial data, disaggregated time data

Examples: Spatial diffusion models.

c) Disaggregated attribute data, disaggregated time data

Examples: Regional econometric models.

Level 4: All data disaggregated in space, attribute, and time

Example: Isard's general theory.

Examination of the level of data requirement of the various concepts illustrates that, for the examples of urban spatial theory given, the availability of data does not affect the conceptual formation of theory.

Level 1 contains some of the more frequently discussed themes of recent quantitative geography, those of rank size rule and gravity theory (principle of least effort). These concepts were, however, formulated between 1929 and 1949.<sup>1</sup> They were taken up again by several workers between 1956 and 1964.<sup>2</sup>

Level 2 contains three categories, each with somewhat different characteristics. The first category contains theories of linear form and concentric form elaborated by workers in the architectural tradition active at the turn of the century. These classic studies include Garnier's concept of the linear form, "Cité industrielle," proposed in 1904, Ebenezer Howard's concept of a concentric form of "Garden City" in 1898, and Le Corbusier's concept of the "Ville radieuse" in 1920. The category also contains work of their modern counterparts<sup>3</sup> in microscale urban design, published between 1964 and 1967, the concepts of central place theory, and the Garin-Lowry models of land use change. The first of these concepts originated over 40 years ago<sup>4</sup> and has been developed more recently (1954-67) by several workers.<sup>5</sup> The

1. Reilly 1929, Singer 1936, Zipf 1949. 2. Carrothers 1956, Berry and Garrison 1959, Berry 1961, Kantorowich 1964, Huff 1964, and others. 3. Alexander 1964, 1967, Alexander and Chermayeff 1966. 4. Christaller 1935. 5. Losch 1954, Berry 1967.

Garin-Lowry models of land use change attempt to use very disaggregated spatial data and are relatively recent developments.<sup>1</sup>

The second category of level 2 contains theories essentially derived from economics and the social sciences. All are data-dependent theories and were developed between 1956 and 1969 by numerous workers.<sup>2</sup>

The third category contains theory characterized by disaggregated time data and generally large volumes of data. Work on migration models by geographers was a development of the period 1957-69.<sup>3</sup>

Level 3 contains three categories, though most of the examples fall into the first which contains concepts employing data disaggregated in attribute and in space. Here again, it can be noted that the availability of data does not affect the conceptual formation of theory. The socio-economic theories were developed between 1925 and 1939.<sup>4</sup> The classic theory of economic rent was first formulated by Ricardo in 1817, and theories of agricultural location and urban location were developed between 1826 and 1926.<sup>5</sup> The main development of these theories and their extensive testing and development, however, occurred between 1961 and 1964.<sup>6</sup> Wilson's<sup>7</sup> technique of entropy maximization should perhaps be regarded as a special case, in that the approach is flexible enough to be used with widely varying levels of disaggregation and types of data. The more frequent applications, however, might be allocated to this category.

The second and third categories at this level both require data disaggregated with respect to time. Coupled with either disaggregated spatial data or disaggregated attribute data, this implies that significant volumes of data must be handled. Work by geographers in this field is a fairly recent development; the spatial diffusion models of Hagerstrand<sup>8</sup> and Pitts<sup>9</sup> were a product of

1. Lowry 1964, 1965, Garin 1966. 2. Isard 1956, 1960, Leontieff and Strout 1963, Webber 1963, 1964, Wilson, A.G. 1967, 1969a, 1969b, Nourse 1968, Wärneryd 1968, Broadbent 1969, Richardson 1969, Forrester 1961, Meier 1962, and others. 3. Hagerstrand 1957, 1965, Bogue 1957, Rogers 1968, Morrill 1965, and others. 4. Burgess 1925, McKenzie 1933, Hoyt 1933, 1939. 5. Hurd 1903, Von Thunen 1826, Haig 1926. 6. Alonso 1964, Chisholm 1961, Hall, 1966, Wingo 1961a, 1961b, Muth 1961, Kain 1962, Herbert and Stevens 1960, Chapin and Weiss 1962a, 1962b, and others. 7. Wilson, A.G. 1969c. 8. Hagerstrand 1965. 9. Pitts 1963.

the early 1960s. Models of regional economic growth<sup>1</sup> are more recent, and some of the very recent work on urban housing such as that under development by the Urban Institute and the National Bureau for Economic Research in Washington, D. C. , would seem to fall into this category.

Level 4 contains the category of theory where the demands for data handling are necessarily high. It is typified by current concepts such as Isard's general theory.

Several suggestions can be made from the above observations. First, and most obvious, is that the conceptual formation of theory is not dependent on the ability to handle data. Secondly, the testing of hypotheses based on such theories by model building received considerable impetus between 1954 and 1967, and to some extent is contemporary with the availability of calculating capacity in computers. Thirdly, the trend in more recent work (1961-69) is toward data-dependent models, which could benefit from increased availability of relevant data. Finally, the development of theory requiring the analysis of data disaggregated in time has not been extensive, and this is possibly due in part to the lack of suitable data and the capacity to handle them.

From the above comments it is clear that the development of geographic information systems will not create theory, but they may speed the course of theory testing and thus allow new avenues to be explored more rapidly. Specifically, they may help theory by making available information that would:

- a) allow current models to be tested over a wider area, or for more disaggregated populations;
- b) allow the development of new models (often derived from old theory) that may incorporate variables not before available in quantitative form (which may, in turn, lead to completely new types of theoretical development);
- c) as a special case of (b) above, provide data on change at places, and the interrelationship between changes at places (that is, the interaction between attribute and time over space), to develop and improve dynamic models of innovation, diffusion, the lagged effects of economic and social change, etc.

1. Richardson 1969.

The above discussion has briefly examined the relationship between formal spatial analysis and the availability of data. The next section will outline the characteristics of the decision-making process in government. Thereafter, the value of existing geographic information systems to formal spatial analysis and decision making will be examined.

#### GOVERNMENT DECISION MAKING RELATED TO DATA USE

The primary function of man is the processing of information. All his decisions are based on his image of reality, his mental model of the real world. Where the decisions (of a government, for example) concern variables distributed in space, the thought processes involved are, to some degree, spatial models. Mental models can be very complex, finely tuned, intuitive processes; they receive data from a variety of sources, fit them into the individual's pragmatic explanation of the courses of action open to him, and form the standpoint from which he (as a Minister of government in Cabinet, for example) must argue his case with his colleagues. The discussion process of Cabinet itself could be likened to an iterative model, receiving variable inputs of information as the debate continues, positions (weighted variables) are changed in debate, and an eventual solution (or lack of one) is reached. It is not suggested that this is a cold, rational process. Any experience of the political or emotional aspects of decision making immediately informs one otherwise. The psychological inputs simply add to the range of information being processed; they are an integral part of the decision-making process and increase the complexity of the inherent modelling exercise. There are no formal surrogates that can adequately replace the mental modelling process in 1974, though obviously both formal and mental models can play a role in decision-making processes. The mental models rely on data presented in such a way that the human mind can visualize and comprehend them. Thus, the role of information systems frequently is to reduce, derive, and reformat data so that they can be understood and incorporated in the mental modelling process (the parallel role for information systems is to supply and manipulate data for input to formal models). Even among governmental institutions, it is easy to envisage a wide range of types of decision-making agency. A typical set of decision-making

institutions is given in Fig. 7.2.

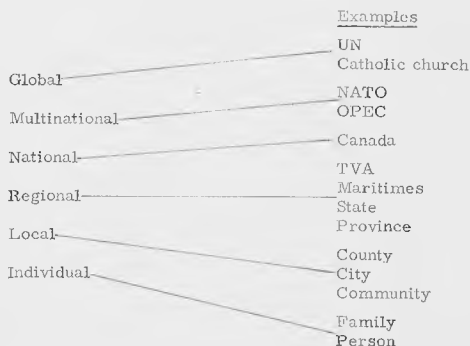


Fig. 7.2. Typical set of decision-making institutions.

The list is arbitrary and other sets of decision-making agencies could be equally valid; one could differentiate between governments in countries at different stages of economic development, for example, or between governments with different political and administrative philosophies. The list above recognizes global, multinational, national, regional, local, and individual levels of decision making. The types of decision-making agencies and the types of decision within institutions that fall into these categories can be readily envisaged. The aim is not to defend this particular set, but merely to indicate the scope of any study of decision making in government. If this set is accepted and only the national, regional, and local forms of government are considered, one can suggest some of the types of decision made by such agencies (Fig. 7.3).

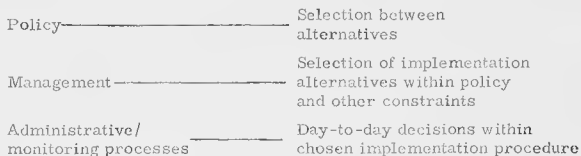


Fig. 7.3. Typical types of decision within national, regional, and local institutions.

The list chosen differentiates between types of decisions in areas of policy, management, and the administrative/monitoring process. The categories are subjective and the actual boundary conditions that separate them are far from distinct; other sets of decision types can be equally valid. Thus, Deutsch<sup>1</sup> reviews several models of political communication and control in government, Bray<sup>2</sup> recognizes general policy reviews, action programs, and operational decisions as separate categories within the British government, and Ward<sup>3</sup> in Canada has discussed government decisions differentiated on fiscal criteria. Again, the aim is not to defend the set presented above, but to indicate the scope of the types of decision that must be considered for future study. In the categories given above, policy decisions are taken to be those where a selection is made between major alternatives of action. Management decisions concern the selection of alternatives for implementing that policy within the constraints of the policy and the jurisdictional constraints of the agency responsible for implementation. Decisions on administrative and monitoring processes are the day-to-day decisions within the chosen implementation procedures.

A hypothetical policy decision at the national level of government, for example, could be to rehabilitate farmers from marginal farms whose income was below the poverty level. The possible jurisdictional, legal, financial, and specific measures would be established. The policy might specify a variety of measures such as farm purchase, formation of farmers' cooperatives, amalgamation of farms into economically viable entities, increased opportunities for extension education in agriculture, direct capital support, subsidies for crop transportation, and so forth. The management decisions related to such policy might be concerned, in the first instance, with identifying farmers on marginal land. This might involve an inventory of marginal land. In a country with the regional diversity of Canada, for example, a management decision may be made, on the basis of various types of data, to employ (promote and encourage) one course of action in one part of the country and another in a different part of the country. Administrative decisions would be at the level of determining whether a farmer was eligible for support under the various categories of

1. Deutsch 1966. 2. Bray 1970. 3. Ward 1962.



the legislation implemented in his area; they would include monitoring the process, measuring the response to the program in the area, and so forth.

The decision-making levels are obviously interdependent and operate within a complex framework within and outside the institutions. A simplistic concept of the relationships between the decision-making levels of national, regional, and local governments in a parliamentary system is illustrated in the diagram (Fig. 7.4) overleaf. The diagram and the comments that follow are as general as possible, but reflect the author's experience with various levels of parliamentary government in Canada.

The hypothetical legislation already outlined will be used as an example to trace the paths followed by the decision-making process as presented in the diagram (Fig. 7.4). The possible contribution of geographic information systems and formal and mental models to the process will also be examined. It must, however, be emphasized that the mixture and value of the various techniques are different for each type of legislation; they vary from year to year, from department to department, and even from individual to individual according to the training and character of key personnel.

Assuming that the diagram represents a national level of government and the policy decision results in parliamentary legislation, the Act will usually be implemented by a single government department, though it may also have an impact on the actions of others. Those involved with the management decision process as defined above are now responsible for choosing the alternatives of action within the legislative framework. For our hypothetical agricultural legislation, the department may first want to gather data, not available from other sources, about the distribution of marginal lands across the country. This step is shown as input from departmental data gathering to the management decision process. A geographic information system may be designed, or an existing one used, to store, manipulate, and generate summaries of the data for the department. Data from other agencies, such as agricultural census data relating to size and condition of farms, gross productivity, average income levels, and so forth, may be derived and reworked to match the data gathered by the implementing agency itself. At this stage, formal

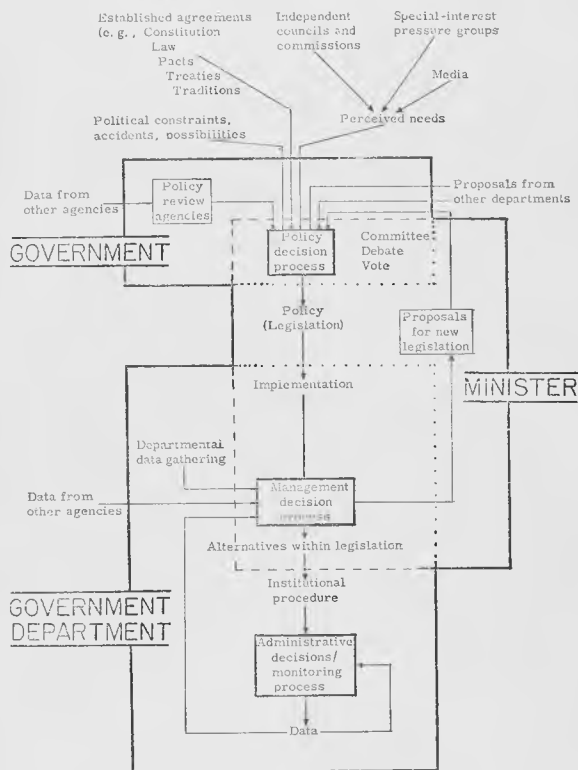


Fig. 7.4. Simplistic concept of policy, management, and administrative decision levels in national, regional, or local government in a parliamentary system.

models may be designed and used. Models may be developed to relate the value of farm land to the distance from markets, to study interregional flows of commodities, to identify regions, to determine which region would benefit from which type of support under the program, and perhaps even to simulate the diffusion of knowledge about the program to the farming community so that its timing and budgeting can be more precise. The management decision process would almost certainly incorporate some such procedures, but would also rely heavily on mental models based on the experience of the resident government agronomists, the provincial ministers of agriculture and their staffs in various parts of the country, and the experience of the departmental administrators themselves.

Within the limits imposed by the legislation, the institutional procedure would be developed and passed to the administrative units for implementation. In the initial stages of administration of such programs, decisions are prescribed and have simply to be implemented. Data are collected as part of the administrative process. However, as difficulties in implementation arise, and as the people in the field find that the program is not accepted or is not having the desired results, their arguments and the supporting data are passed back to the management decision process. Such arguments frequently involve logical, well-documented, data-dependent reports, but in our hypothetical example they would almost entirely be based on the mental models of the experienced field officers.

At this stage the processes of management and administrative decision can be considered to act together as an iterative procedure. Fresh data from administrative sources, perhaps additional data from other agencies, representations to the minister of the department, and his redirection of priorities or realignment of procedural emphasis are some of the factors that change.

Programs bloom, wither, and die in time, as the conditions in which they operate change, as the need for them becomes less real or apparent, or because the legislation originally passed down does not allow the program to succeed. In our hypothetical case, the last situation might occur if the original legislation had been framed to give aid directly to farmers on the

marginal farms of a certain rural area, without recognizing that their plight might be inextricably bound to the total economic system in the rural-urban region that included their area. It might become clear as the program proceeded that the main effect of the policy as implemented was a rapid increase in rural-to-urban migration, and that the "rehabilitated", displaced marginal farmers were becoming urban welfare recipients and creating a more severe social and economic problem than the one initially faced.

In such circumstances, it is common for a department to prepare proposals for new legislation to be submitted to government. These may either agree or conflict with proposals for new legislation from other departments. A Health and Welfare agency, for example, may have ideas on the appropriate answer to our hypothetical problem completely different from those of an Agricultural agency. The policy decision process must be responsive to many inputs. The primary link between the levels of management decision and policy decision is the function of the elected minister. His awareness of the need for a change in policy comes not only from the senior civil servants in his own department (or from their negotiations with their counterparts in other departments), but from the advice of independent councils and commissions, from special-interest groups such as farmers' associations, from personal contacts, and from the media. Commissions, such as the Economic Council of Canada, may well use formal models in preparing a submission (to the minister, to parliament, or for general publication) on the state of the farming economy. To a lesser extent, agencies such as farmers' associations employ economists to help them prepare their arguments. It is probable, though, that most submissions are based on mental models and, in the case of the media, they are perhaps limited to the contribution of data and opinion of varying degrees of reliability.

The process of policy decision in government works in the milieu of established agreements; it operates within the constitution (which may have a particularly strong influence on policy under a federal system), the law, existing treaties, international tariff agreements, and traditions of inter-governmental consultation. The process also works within political constraints, accidents, and possibilities; "political" here is used in a

broad sense to include the interactions with the business, labour, and religious communities. These variables must all be considered in policy decisions. Policy frequently is influenced by background studies, and proposed policies are often subject to review, by agencies such as Treasury Board, Department of Finance, and Privy Council Office. These agencies may or may not use formal models in their evaluation of policy, but a strong tradition of economic input/output models and national accounts (and to some extent regional accounts) underlies major fiscal decisions. However, the advice specific to a particular policy will certainly incorporate a strong element based on mental modelling.

The parliamentary process of policy decision by committee discussion, debate, and vote could be described as being based on mental modelling. It receives inputs of a wide variety of data, some of which may come from geographic information systems in the form of maps, diagrams, statistical summaries, flow charts, and the like. From this process emerges (it is hoped) new policy which, at national level, may be translated into legislation.

From the above simplified discussion of the interactions among decision-making levels of government, it is clear that a mixture of formal models and mental models is involved in decision making. In general, formal models probably make their major contribution at the management level, and as an element in the processes of policy review (both background review and proposed policy review). All three levels of decision making, however, are dominated by the use of mental models. Decisions at the level of the administrative/monitoring process are dependent on mental models because, to a great extent, actions at that level are prescribed by institutional procedures (decision models are essentially provided). Although data gathering (program monitoring) is a fundamental activity at this level, few resources are provided for abstract thought. The valuable advice from the administrative level relies heavily on the mental modelling processes of experienced administrators. Decision making at the management level, in contrast, has the resources and the need to use formal models (there is usually a "Policy and Planning" division in government departments at the national level in Canada). Nevertheless, the inputs at this level are

numerous and not necessarily quantifiable. Their integration relies upon the experience (mental models) of the senior administrators and, in particular, on the interaction between the department (civil service) and the minister (elected official). Despite the input of formal models to the policy review process, policy decisions must also be considered to be dependent on mental models. The decisions are based on a wide variety of factors, of which many are essentially quantitative, some uncertain, many dynamic, and all, together, invariably complex. Decision theory and, in particular, the recent development of theory on decisions under conditions of uncertainty are not developed (in 1974) to the stage where the overall policy process of government is amenable to formal models.

With the above discussions of the characteristics of formal spatial analysis and an outline of the decision-making levels of government in mind, the contribution of geographic information systems to the overall situation will be examined.

#### CONTRIBUTION OF DIFFERENT TYPES OF EXISTING GEOGRAPHIC INFORMATION SYSTEM TO FORMAL SPATIAL ANALYSIS AND TO DECISION MAKING IN GOVERNMENT

The five broad categories of geographic information systems, based on differences in the mixture of their characteristics of manipulation capability, capability to handle volumes of data, and resolution of location identifier, were recognized in Chapter 6. They are summarized in the five rows of Figure 7.5.

The first category identified is that of systems whose primary function is image handling. They are not designed to handle complex sets of data related to the location identifiers, but exist to produce images, to make the production of maps more rapid or more economical, or to facilitate visual display. Their usefulness can be inferred from their end products.

Those with a storage format of arbitrary regular areas (Type 2), the SYMAP-MIADS group, are used to generate inexpensive, simple, visual expressions of statistics. Typical products<sup>1</sup> are graphic illustrations of income density in Southern New England; of distribution of sediment classes on a portion of the Ohio River floodplain; of industrial employment as a percentage of total

1. Pictorial illustrations of the examples cited are contained in Warntz 1971 and Tomlinson, R. F. 1972.

Group characteristics	Typical systems	Primary utility	Primary use
Primarily image-handling systems: Field of automatic cartography Varied location identifiers	SYMAP, MIADS, etc. AUTOMAP CAN, HYDROGRAPHIC	Facilitate visual display Provide economic/rapid map production	Mental models
Relatively high data volume: Low-order location identifiers and manipulation capability	CENSUS LUNR, MINN	Generalize large volumes of data	Mental models Formal models using aggregated data
Medium manipulation capability: Medium to high data volume Medium-order location identifiers	FRIS, GRDSR DIME, SACS, OEM NORDPLAN WWW  INSIGHT	Developed by, and reflect needs of, census or administrative agencies. Attempt to handle data volume with greater facility than manual techniques or systems using low-level location identifiers.	Wide potential application to formal models
High manipulation capability: Limited data volume Medium to high-order location identifiers	PIOS, NRIS MAP/MODEL GLIMMS	Facilitate graphic tasks that are prohibitively expensive by manual methods	Mental models Source of derived data for formal models
High-order location identifier: High manipulation capability Low to (potentially) high data volume	CGIS	As above Nation-wide data base structure	Mental models Source of extensive derived data for formal models Potential application to formal models

Fig. 7.5 Categories of geographic information systems, their characteristics, primary utility, and primary use.

employment; and percentages of land in residential use (separate maps, same area) in the Northeastern Corridor Transportation Project, U. S. A.; of air pollution from 2-hour sulphur dioxide readings in St. Louis; of change in population by census district in Canada; or of the percentage of coconut-growing acreage suffering from blight in Montegna Bay. Many hundreds of similar products could be listed. The process presents spatial statistics in a form that can be more easily comprehended by an observer, and is used to illustrate reports, to support a verbal presentation of a concept or point of view, and (occasionally) to trigger new understanding of distributions. Essentially the systems provide input to mental models. The AUTOMAP system can rapidly (in 2 hours) produce a simple outline map of any place in the world, at any scale (from 1:10,000 to 1:1,000,000,000) on any one of 17 map projections. Pertinent data are annotated by hand on the computer-produced outline map. The maps are provided by the Central Intelligence Agency<sup>1</sup> of the U. S. A. as a background for daily briefings of the Director of C. I. A. and the President of the United States. The Canadian Hydrographic System provides a means of rapidly producing (revising) hydrographic charts from new soundings and hydrological observations. It is typical of systems of "automatic cartography" that are in early stages of development in traditional map-making institutions.

The overall capability of these systems is to reformat data flexibly so that they fit human mental understanding of space, and in one sense (that of generalization) to reduce data volume. They also can produce innumerable graphic images, which must subsequently be visually examined. These are, of course, the characteristics of traditional cartography. In all cases, the systems in this category make their primary contribution to mental models. They are conceptual extensions of manual cartographic practice. Their value lies in the degree to which they can lower the cost or increase the rapidity with which graphic images are produced.

The second category contains two apparently dissimilar types of system, the traditional method of processing census data and the systems with a storage format (large grid) of arbitrary regular areas (Type 2). The underlying functions, however, are the same: they both generalize large volumes of data

1. Schmidt, W. 1973.



and both have arisen from the need of national or regional governments to gather statistics about their domains and subsequently summarize the data for human comprehension.

The value of such systems at the administrative level of national, regional, and local governments is marginal. Because they summarize data into relatively large areal units for reasons of convenience or confidentiality, they effectively obscure the data on which most administrative decisions are made. This is one of the reasons why administrative levels of government tend to gather their own data. This situation is only ameliorated when the summary area coincides with an administrative area. Even this does not provide the disaggregated data relevant to decisions at the administrative level.

The usefulness of such systems to the management and policy levels of decision making rests on total coverage of the area that the decision concerns. The value of summaries of census data to national and regional levels of government (while subject to major improvement) is not in question. The total coverage of the governmental domain is inherent in the census process; the principal summaries follow one (usually political) hierarchical set of major administrative areas (though there is nothing "natural" about such units).<sup>1</sup> If, however, similar large-grid-type coverage of the decision-making area of national or regional governments is employed, the advantage of regular areas is gained but the manual process of assigning data from other graphic formats to grid squares is enormous. The resulting grid squares are not aligned to administrative boundaries, though these may be approximated if some further data loss is acceptable. They can, however, present the generalized data in computer-generated displays of the SYMAP type.

Such systems are useful only to formal spatial models that use spatially aggregated statistics. The census has the merit of measuring many attributes, and does at least provide a wide variety of spatially aggregated statistics. The design of the current data banks with a large grid system assumes that the required statistics are the aggregated ones (but there has been little contact to date between system designers and spatial analysts), and that the size and shape of the arbitrary unit of the system are appropriate to the level of

1. Several discussions have considered the problems inherent in using data from administrative units that are not similar or comparable with regard to the data aggregated within them. See, for example, Harvey 1969, Chapter 19, Robinson 1956, Thomas, E. N. and Anderson 1965, and Duncan et al. 1961.

aggregation of the model concerned.

For all types of system in this category, only limited operations of manipulation can be carried out on the summarized data. This is not a computer limitation, but is inherent in the nature of summarized data. Grouping and display are straightforward, but significant measurements and comparisons can be made within the data set only when very large areas are involved (and the Type 2 storage format is thereby approached).

Perhaps the real value of this approach to spatial data handling is the basic fact that internally consistent data sets for national and regional areas are actually gathered. The familiarity of census statistics obscures this fact, but it is more apparent in the recently developed large grid systems. The LUNR<sup>1</sup> system of New York State, for example, required aerial photography of the entire state (140,000 km<sup>2</sup>) and state-wide identification and classification of land use categories, recreational facilities, transportation facilities, water resources, and so forth. Although these data were subsequently summarized into large grid squares, the state has wisely made the original data available as well as the summary data.<sup>2</sup> The demand for the basic data in graphic form has so far outstripped the demand for the computer-stored, summarized data. The graphic basic data has primarily been used for decision making at the administrative level of government. The "system" has thus made more data observations available, even though it has essentially depended on manual effort to achieve this. The author knows of no instance where the summary data in large grid format from that particular system has been used in a formal spatial model; the grid displays have been of some value in the mental modelling process at management and decision-making levels of state government.

The third category comprises systems that assign data to point locations or to line segments, and attempt to bring additional capability for spatial manipulation to large volumes of data. It is tempting to say that no one is more aware of the manipulative limitations imposed by large data volumes and low-level location identifiers than the census and administrative agencies themselves. Whether or not this is true, the primary development of the

1. See Chapter 4, pp. 99-103. 2. Remember that the non-graphic store does not contain the same data set as the graphic store. It contains a selected set of measurements, manually derived from the graphic store.

GRDSR, DIME, FRIS, OEM, WWW group of systems has been undertaken by agencies responsible for data gathering rather than implementation of programs. There are some exceptions. Various cities in the United Kingdom and North America (for example, Coventry,<sup>1</sup> Ottawa,<sup>2</sup> and Detroit<sup>3</sup>) have experimented with point-geocoded parcel files for administrative purposes. The proposed<sup>4</sup> point referencing system in Great Britain stems from the interaction of planning departments at the local government level and the (now) Department of the Environment. The meteorological services of various countries participating in the WWW system collect the data for their own analysis (though that probably should not be thought of as the end use of the information). However, the pattern is clear; the initiative has been based on pragmatic considerations rather than intellectual curiosity.

The limitations of this category of system (information reduction with use of non-optimal location identifiers, restricted inherent connectivities of point data sets, and so forth) have been discussed in the text.<sup>5</sup> These limitations are minimized where closely spaced location identifiers are employed. A common use of the systems is to handle urban data.

In spite of their limitations, it can be said that such systems allow access to, and spatial manipulation of, data that would otherwise be unavailable, because of the cost of handling high volumes or the submergence of disaggregated data in large area summaries. They provide potential access to many measured variables in census and administrative records that previously could not (in practical terms) be spatially related to one another or even spatially grouped together.

With the exception of weather monitoring (WWW), these systems have been developed and tested only since 1968. Apart from WWW, the rural capability of GRDSR, and the contribution of INSIGHT (which will be discussed separately below), the systems handle urban data. Owing in part to their recent development, their implementation and use in decision making have just begun and are potential rather than actual. Several examples can, however, be given to illustrate the capability that exists.

1. Coventry City Council 1971. 2. Symons 1972b. 3. Becker 1971.

4. Department of the Environment, U.K. 1973. 5. See Chapter 4, pp. 95-96.

At the administrative level of decision making, the police department in Boston,<sup>1</sup> U. S. A., uses its DIME system to dispatch emergency cars by address, intersection, and street location. Police "incidents" are geocoded and police patrol routes are revised on the basis of the data. Los Angeles<sup>2</sup> and other cities have used census data assigned to point coordinates to develop tests of potential carpool candidates on the basis of proximity to residences, work hours, and distance to work. The daily route maps for solid-waste collectors in Baton Rouge, Memphis, Shreveport, and other cities in the U. S. A. are revised by use of a line segment system (RAGS)<sup>3</sup> similar to DIME; collection costs have been reduced by 10% to 25%, daily routes have been reduced from 200 to 175, and from 65 to 32, among other results. Similar applications include assignment of pupils to schools<sup>4</sup> and creation of street maintenance and repair schedules.<sup>5</sup> These processes primarily rely on specific data as input to mental models, but occasionally they employ simple partial models for processes of network optimization.

At the management level of decision making, the location of facilities related to disaggregated population characteristics is a typical example of decisions that are made between alternatives within a policy. Decisions have been made on location of hospitals,<sup>6</sup> fallout shelters,<sup>7</sup> new schools,<sup>8</sup> unemployment or welfare offices,<sup>9</sup> and commercial establishments,<sup>10</sup> and alternative highway routes<sup>11</sup> have been evaluated. Similar choices between alternatives have been examined in the evaluation of modal choice<sup>12</sup> between urban transportation networks, given a policy of increasing commuting service. The data are used in simple spatial models in most instances.

At the policy level of decision making, data have been collected to display graphically the high rates of abandonment of subsidized housing;<sup>13</sup> to demonstrate the impact of a new rapid transit system<sup>14</sup> on population, land use, and public safety around new stations; to determine the potential demand for bank services;<sup>15</sup> and to provide background data for school desegregation policies.<sup>16</sup> Basically, the data provide input to mental models.

1. Krasnakevitch 1973. 2. Hoelzel 1973. 3. Owen 1972. 4. Salomonsson 1973.
5. City of Vancouver 1965. 6. Dowd 1973. 7. Treichel 1973. 8. Barrett 1973.
9. Westerfield 1973. 10. Healy 1972. 11. Krasnakevitch 1973. 12. Barrett 1973. 13. Farnsworth 1973. 14. Barrett 1973. 15. Cooke 1973.
16. Almendinger 1973.

Data from this category of system are suitable for many types of more complex formal spatial modelling. They have been used for input to gravity models,<sup>1</sup> forecasts of trip generation models,<sup>2</sup> facility demand models (the CRAM<sup>3</sup> model was briefly described in Chapter 4, pp. 120 - 121). Others can easily be envisaged.

The potential value of this type of geographic information system for urban decision making and formal spatial analysis is clearly high. Network (line segment) systems probably are somewhat more potentially useful than point data systems, principally because the connectivity inherent in their file structure is more specific. The conversion of the point data system FRIS to a line segment system by NORDPLAN illustrates this point of view. Topologically rigorous systems would seem somewhat more useful than systems with strictly metric coding, if only because of the stronger internal editing possible in the former and the practical problems associated with digitization in the latter. Interactive systems can be expected to add a higher dimension of manipulation capability, and of usefulness.

INSIGHT<sup>4</sup> (discussed in Chapter 5, pp. 175-178) is typical of a new breed of system, those with interactive capability. It falls into the third category because it can provide access to large data sets, has a medium level of capability for spatial manipulation, and can use medium-order location identifiers (essentially point data sets in INSIGHT). A similar interactive system operating on line segment data has been developed as an adjunct to SACS by the Urban Data Center in Seattle. The application of this approach to the large data sets already geocoded (such as GRDSR and DIME) would add a capability for heuristic browsing and comparison to already useful systems. It would also add a valuable new tool to the process of mental modelling, which in turn might open lines of inquiry leading to the formulation of hypotheses that could be tested by noninteractive use of the same data bank. Again, it can be noted that the capability was developed in response to the need of a data gathering and administrative agency of government.

The fourth category contains the "coarse" polygon systems (MAP/MODEL, PIOS, NRIS, GIMMS, etc.). They are characterized by a relatively high capacity for manipulating their spatial data; they all carry out overlay with

1. White, M. S. 1973. 2. Barrett 1973. 3. Farnsworth 1970. 4. Phillips, R. L. 1972. 5. Calkins 1973.

some degree of success, and can manipulate, compare, and derive a variety of spatial measurements from their stored data. "Coarse" polygons are medium- to high-order location identifiers, though, as yet, the systems handle low volumes of data related to them. In essence, they are surface-descriptive systems, designed to facilitate graphic tasks that are prohibitively expensive by manual methods. In the main, their development has resulted from the work of geographers and planners, who, as a group, are professionally interested in surface characteristics of specific areas and are aware of the effort involved in handling graphic data manually.

In their current state of development, none of the systems is of immediate benefit to decision making at the administrative level. Given a substantial increase in the areal extent and volume of their stored data, and increased ease of use, they could become useful for monitoring change, or as a readily updated data base against which to check compliance to regulations, such as zoning decisions or building permit applications. Their main use, however, is not anticipated at the administrative level.

Their primary value lies in their contribution to management and policy decisions. They can compare (overlay) two or more attributes of surface-descriptive data, such as potential land use with actual land use, forest categories with slope categories, soil data with traffic zones (PIOS),<sup>1</sup> or compare maps of a single attribute prepared at different times, such as land use of the same area in 1965 and 1970. This process is analogous to "sieve" mapping and allows, for example, several constraints on land development to be combined and areas suitable for development to be pinpointed, or areas with multiple specified characteristics to be identified for facility location. A very interesting case of the approach being used to define legally an established community centre in Romulus, Michigan, in terms of overlapping social services and a range of central place functions was recently reported by Kolars and Nystuen.<sup>2</sup> In all cases the systems have the ability to produce graphic images (maps) showing the results of their operations. It must be noted that much of the original surface-descriptive source data comes only in the form of maps (or similar graphics) using irregular polygons. Visual comparison between graphic images not only is

1. See discussion of PIOS in Ch. 5, pp. 144-152. 2. Kolars and Nystuen 1974.

time consuming, but can be misleading.<sup>1</sup> The overall value of this category of systems lies in making this type of derived data available to the decision-making process; such systems can increase the number of observations that relate to a decision.

The systems can derive quantitative measurements of the correlation, generate aggregated data within newly defined, irregular-shaped areas, and derive quantitative measurements of numbers of attributes, lengths of lines, and extent of areas from their data base. Thus, they can increase the number of derived observations that are economically available for input to formal spatial models. (PIOS was, in fact, specifically developed to provide data for a model of urban development.)

The fifth category is characterized by high-order location identifiers (fine polygons), but equally importantly by high capability for manipulation and the potential capability for handling large volumes of data by a file structure already designed to accommodate them. An essential aspect of the category is that a basic file structure for spatial data allows region-wide or nation-wide data to be stored and efficiently retrieved. This implies that the inherent problems of internal transformation between coordinate systems, edge matching of source map sheets, data compaction, and efficient file structures for spatial data storage and retrieval have been overcome. At the present time (1974), it also implies that an efficient process of graphic to non-graphic conversion has been devised.

The underlying intent of systems in category 5 is to use non-graphic storage as an alternative to graphic storage, to serve as a permanent source of spatial data otherwise available only on maps. Its objective is to provide economical access to the substantial quantities of source data describing attributes of the earth's surface recorded only on maps. Although this objective is a simple conceptual extension of the capabilities of systems described in category 4, it is, in many ways, parallel to that of systems like GRDSR and DIME (category 2) that not only assign location identifiers to large data files, but more importantly allow computer-aided spatial manipulations to be performed on such large data sets. The previous constraint of such census data sets was that no disaggregated location

1. Ekman et al. 1963.

identifiers were available for otherwise machine-readable data. The previous constraint of map data sources was that their disaggregated location identifiers were not in machine-readable form.

CGIS<sup>1</sup> is an example of a system in category 5. Its limitations are numerous. It handles data only in irregular polygon formats (although the capability for handling point-referenced data is now being added). Despite the fact that it is designed as a store for large amounts of data, there is no browsing capability; all processing is done by batch-processing. Its manipulation capabilities do not include handling irregular polygons that describe continuous, single-valued functions; that is, it cannot manipulate or generate contours.

The value of the type of systems in category 5 follows a similar pattern to category 4 systems. In their present form they can make little contribution to decision making at the administrative level, although with considerable further development, and particularly an interactive man/machine interface, they might approach the usefulness of category 3 systems for decision making at that level. The question concerns the need for surface-descriptive data in such decision making rather than the requirement for techniques of spatial data handling. The primary value of category 5 systems is at the management and policy-making levels of decision making.

CGIS was designed in a government department<sup>2</sup> in response to the needs of that department to gather and analyse data for management decisions. The department had undertaken to take an inventory<sup>3</sup> of the resources of the productive areas of Canada in the categories of present land use, agricultural capability, forest capability, recreational suitability, and suitability for wildlife. Each of these attributes was separately observed (as each required separate observational skills) and encoded on separate series of map sheets. All data were mapped in the form of irregular polygons. Between 10,000 and 20,000 manuscript map sheets were initially involved. The decisions to be made by the department concerned choices among competing proposed land uses, based on comparison of the values of the various land attributes; that is, it was necessary to compare the attributes by overlay of the map sheets. At this management level of decision making, the retention of boundaries

1. See Ch. 5, pp. 152-168 and Appendix. 2. Department of Forestry and Rural Development, Government of Canada. 3. The Canada Land Inventory.



was significant. It was also desired to overlay the data on land attributes with the boundaries of census enumeration areas so that the socio-economic attributes of areas could be similarly examined and correlated. In addition, summaries of widespread correlations between the variables were needed as a basis for policy decisions.

These tasks required high-order (but conceptually straightforward) manual cartographic techniques, which were prohibitively expensive in terms of effort and time. The contribution of the CGIS in carrying out these techniques was to provide data to the mental modelling process at management and policy levels of decision making. Although the system must still be considered to be experimental, it has provided test data as a basis for mental models at management and policy levels for comprehensive development programs on land use in Prince Edward Island, Nova Scotia, the Gaspé Peninsula, and Manitoba.

The capability exists within the CGIS to derive measurements and comparisons from the stored data and produce the results in quantitative terms, to assign any result to any point location, or automatically to assign weighted attribute values to any size of overlaid grid square (or other arbitrary regular area). It thus has a high potential as a source of derived data for input into formal spatial models. The file structure of the system can accept any descriptor of an attribute and so can respond to as yet unspecified inputs to formal models that relate to surface attributes within irregular polygons. The author, however, knows of no instance where data from the system have been used in a formal spatial model.

#### OVERALL VALUE OF EXISTING GEOGRAPHIC INFORMATION SYSTEMS

In considering the value of geographic information systems, it is worth noting the difference between a store of information and an information system. Earlier in the thesis<sup>1</sup> it was stated that the usefulness of a store of information depends on whether the type (attributes), resolution (level of disaggregation), and format (type of location identifier) of data are relevant to the decisions (analysis) being made, and also on the ease with which the data in the store (in terms of their attributes, volume, and format) can be manipulated to produce the information on which decisions

1. Ch. 4, p. 85.

(analysis) are made. In view of the discussions above, it seems appropriate to add that the value of the store of information also depends on whether its contents are perceived to be needed for explanation or decision, that is, whether the training of the users has provided them with formal or mental tools that use such data.

The value of geographic information systems must be considered within this concept of the usefulness of stores of information. It is clear from the discussion above that the current contribution of geographic information systems to both formal and mental types of spatial analysis is twofold: they increase the amount and variety of existing data that can economically be available for spatial analysis, and they increase the capability of manipulating those data to produce the information appropriate to specific types of analysis, that is, where characteristics of place are involved. It has been shown that these functions rest on increases in the capabilities for reading data, assigning them to location identifiers for subsequent spatial analysis, deriving data by measurement and comparison, and generating graphic displays.

## CHAPTER 8 - FUTURE DEVELOPMENT OF GEOGRAPHIC INFORMATION SYSTEMS

### DEVELOPMENT OF SYSTEM CAPABILITIES ASSUMING PRESENT COMPUTER CHARACTERISTICS

Several useful lines of future development can first be proposed within the deliberately narrow view of the value of geographic information systems themselves. Assuming present (1974) characteristics, capabilities and costs of computers, it is possible to envisage development of both the available data base and the manipulative ability underlying spatial analysis and decision making.

#### Improvements of the Non-graphic Data Base

The trend in improvement of techniques for transformation from graphic to non-graphic format has already been noted.<sup>1</sup> Certainly this is at present a major constraint on the availability of spatial data that can be processed by computer. The nature of the new techniques in this field has already been examined,<sup>1</sup> but it must be recognized (though it is frequently overlooked) that the problem is not simply one of "digitization", but rather of transforming an error-prone manuscript to an acceptably error-free non-graphic file. This end product is particularly significant in the current stage of development of computers themselves, as they are not selectively "forgiving" in their subsequent analysis of data; they may be efficient but they cannot intelligently ignore nonlogical errors. The interactive capability that couples a man's mind directly to a computer, and the sophisticated development of computer programs with a capability for multiple, heuristic, internal editing, are still in a most elemental state. Consequently, the techniques of pre-editing graphic data for digitization and correcting errors in the subsequent file structure are essential steps and costs in the process of transformation from graphic to non-graphic format. Improvements in these aspects of the process are just as important as improvements in digitization, and must receive more attention than at present.

Increases in the store of machine-readable data potentially applicable to a specific problem can come from the encouragement, by persuasion or legislation, of compatibility between the attributes, location identifiers, and time

1. Chapter 5, pp. 171-175.

identifiers of data. The case for data standardization has been frequently stated. The case for national standards of location identifiers has been recently examined<sup>1</sup> in the U. S. A. , and even the national savings on costs of such a step were estimated. Undoubtedly such long-term objectives are desirable, but as an interim measure compatibility, rather than standardization, may be more readily realized. Compatibility can be encouraged through the establishment of guidelines (as recently attempted in the U. K. ,<sup>2</sup> but not in many other countries), by the advice of centrally funded sources of expertise available at no cost to participating levels of government (an approach that has application in a federal system), by design requirements attached to the funding of programs that involve data gathering, and by agreement between agencies with mutually supportive programs. With the exception of the last example, data compatibility at this level is considered to be a technical issue which can be resolved by those concerned with the design of data handling systems. Data compatibility does not guarantee the availability or exchange of information, but it would make possible many developments that are now impossible. It would turn the matter into one of legislation and not of physical frustration.

Increases in the store of non-graphic spatial data can come from the development of techniques that accept non-graphic data directly from quantitative data gathering processes (sensing, surveying, and so forth). It is paradoxical that surveys for mapping purposes often gather data first in numerical form. These data are then turned into graphic data, which must be transformed to a non-graphic form again for computer storage. There are, of course, advantages in the traditional process. The most significant is human interpretation of the original numerical data to produce a graphic image. It is important, though, that the concept of sources for non-graphic storage is pushed back beyond the graphic image, for two reasons. The first is the fact that in traditional "interpretation", numerical source data are frequently replaced by graphic material as the only available "source". Contours are presented as the source data of a surface instead of the measured point observations on which they were based. Delineations of soil types replace the many quantitative measures of soil characteristics gathered with considerable effort at each sampling location. The basic quantitative data are

1. Barraclough 1972. 2. Department of Environment, U. K. 1973.

effectively lost, but could be retained and retrieved in concert with the interpreted image from adequate non-graphic storage. Secondly, sensors are becoming available that record attributes of the earth's surface only in the form of large quantities of numerical codes. The ERTS-1 satellite is an adequate example of this trend.<sup>1</sup> The techniques for accepting these data directly into non-graphic storage so that they can be reliably compared with other data are in an even more embryonic form than spatial data systems themselves. But they are being developed, and the further development of geographic information systems to handle such data will certainly add to the volume, coverage, and timeliness with which specific attributes of earth-surface data are stored for use.

#### Improvements in Manipulative Capability

Increases in the ability of geographic information systems to perform manipulations can stem from several lines of development. Two of these have been mentioned in the text. The first is the further development of interactive capability.<sup>2</sup> This is a most significant addition, not only because it may be of considerable benefit to the economics of error correction, but also because it allows the user to "see" what is otherwise an invisible data bank. The singular advantages of graphic forms of storage are the speed of access to data within the store and the capability of browsing heuristically through a map collection. This capability is lacking in most current geographic information systems, and it is an important failing. Interactive systems are now becoming feasible, primarily because of the advent of the mini-computer and low-cost cathode ray tube displays (the former can act as a communications device to a larger computer if additional capacity is called for, as in the INSIGHT<sup>3</sup> system). Efficient file structures are needed for such systems. Considerable effort is needed to develop interactive manipulation capability (in addition to browsing and retrieving of data). The interactive approach, however, is an important key to the usefulness of non-graphic data storage for the larger data sets that can be expected.

The second line of development in manipulation mentioned in the text<sup>4</sup> was that of multi-format systems for handling spatial data. This capability has not been widely developed to date, partly because of the attractive simplicity

1. See Chapter 5, pp. 178-179. 2. See Chapter 5, p. 175. 3. Phillips, R.L. 1972; see also Chapter 5, pp. 176-179. 4. See Chapter 5, p. 180.

of single-format systems for primary development efforts, and partly because of the lack of specified need to relate different spatial formats in a more efficient suite of manipulations. Undoubtedly the attributes of the real world can be more accurately represented in any storage system if spatial formats can be mixed. Equally, it is clear that certain forms of spatial analysis can be carried out more efficiently if data are available in matching formats (matching each other and matching the techniques employed in the spatial analysis). The possible loss of information incurred by changing the format may not be significant to the problem being solved. This capability is being explored in various systems,<sup>1</sup> but if the requirement for manipulation can be specified adequately, it represents a fruitful line of development.

Improvements in the capability for manipulation can also come from greater statistical understanding of the manipulation processes themselves. Again, a paradoxical situation presents itself. In the first instance, many potentially useful procedures of manipulation, which systems are already inherently capable of performing, are not attempted.<sup>2</sup> In the second instance, manipulations are carried out and the results are accepted without the assumptions underlying the operations being examined.

Most of the systems in category 4 or 5 given in Chapter 7 (pp. 247-252) have the capability for overlay (patch on patch). Measures of association of the comparison can be generated. It is doubted, however, whether the inherent possibility of producing spurious results has been examined by system designers. There is even less evidence of the design of manipulation functions that could investigate such comparisons, for instance, by cross-correlation between various scales obtained by spectral analysis, though this would be a fairly straightforward procedure to implement even with the existing file structures. Clearly, statistical methods must be integrated

1. Chapter 5, pp. 183-184. 2. The reason for not carrying out manipulations that systems can perform is complex. Partly it is lack of a specified need, partly it is the inability of non-user designers of systems to recognize the capabilities of their system, partly it is the fear attached to any state-of-the-art project and expressed by the attitude "Let well enough alone," and partly it is lack of resources ("The designers are having a hard enough time making the system work as it is, without wanting further tasks thrust upon them.") Essentially, it is a reflection of the early state of development of the systems.

with the capability for manipulation in system design if sensible use is to be made of existing and future geographic information systems.

The importance of this line of development is emphasized by an example, quoted by Dangermond,<sup>1</sup> of the second instance: where a manipulation is commonly carried out without examination of the assumptions underlying the operation. The case is the simple one of areal overlay (intersection of sets) where it is assumed that the polygons being overlaid are homogeneous within themselves. The Commonwealth Scientific and Industrial Research Organization (CSIRO) in Australia manually overlaid three sets of polygons and carried out field checks to determine whether the attributes that coincided at specific points on the maps actually coincided in the field. Only in 38% of the cases could they correctly predict the presence of the three attributes together on the ground. This isolated result may be due to a variety of factors, including the possibility of inaccurate original survey.<sup>2</sup> It does, however, illustrate the lack of understanding of even straightforward processes of spatial manipulation which are regularly carried out. The problem extends to the use of standard statistical procedures. A common assumption in statistics is that variables are independent, hence the difficulty of applying them to time series or spatial data. The problems of autocorrelation have, of course, been recognized,<sup>3</sup> but the fact remains that very few standard statistical procedures are aspatial.

In general, statisticians have not developed techniques that are readily applied to spatial prediction. There are exceptions. Weiner<sup>4</sup> in 1930 introduced work on general harmonic analysis, and Matheron<sup>5</sup> built on this work and introduced a new field of statistics, different from the traditional one in that it assumes that variables are related to each other. Some geographers<sup>6</sup> are starting to apply these approaches to spatial analysis, but such sophistication is not common. More frequently, an attempt is made to convert spatial data to a form that fits the underlying assumptions, or even to proceed without understanding the assumptions. Such manipulations may acquire even more spurious validity when they are the result of a supposedly accurate computer procedure. Improvement of capabilities for manipulation must rest on a better understanding of the statistical basis of the operations

1. Dangermond 1973. 2. Which argues strongly for the development of indices of the reliability of source data and their incorporation as flags in non-graphic storage. 3. Cliff and Ord 1973. 4. Weiner 1930, 1958. 5. Matheron 1970. 6. Switzer, P. 1973, Rayner 1974, for example.

involved.

A vital line of development, related to the problems mentioned above but considerably more fundamental in nature, is investigation of the logical properties of the languages available for ordering spatially distributed data. Currently, two different types of language are used for identifying geographical individuals. One language type is used to describe the attributes, and the other the space-time properties, of the individual (the "substance language" and the "space-time language" referred to by Wilson<sup>1</sup>). Notions of geographical similarity of individuals thus make it imperative to describe sameness in at least two languages. To derive individuals specified in one language (say the space-time language) from individuals specified in the other language, an adequate translation process is required. Harvey<sup>2</sup> suggests that the mixing of such languages underlies the difficulty surrounding regional concepts in geographic thought. The extension of geographical problems into four-dimensional space (a very simple concept) immediately illustrates the limitations and awkward structures resulting from the use of current languages. An example, provided by Flowerdew,<sup>3</sup> would be the description in a substance language of the distribution in a country of the "working class"; such descriptions cannot be efficiently handled in existing geographic information systems. Although it would be valuable, it is not possible in this thesis to review or initiate discussion on geographic concepts of space, the characteristics of existing languages, and the needed types of development. Important recent contributions to other than existing languages have been made: by Hägerstrand<sup>4</sup> and Grigg<sup>5</sup> concerning the concept of the "geographical individual"; by Tobler<sup>6</sup> who extends the everyday concept of geographical space beyond the standard Euclidean framework; and by Dacey,<sup>7</sup> Flowerdew,<sup>8</sup> and Betak<sup>9</sup> in their investigations of the concepts and further development of spatial language. Despite these examples, geographers have done little work in this critical field. Clearly, the facility with which one can describe the spatial interaction of surface attributes depends on the language used for the description. The design of file structures for geographic information systems also depends on that language. The efficient design of geographic information systems, and in particular their provision of data

1. Wilson, N. L. 1955. 2. Harvey 1969, Ch. 14, p. 216. 3. Flowerdew 1974.  
4. Hägerstrand 1953, 1963, 1967b. 5. Grigg 1965, 1967. 6. Tobler 1963, 1964b, 1970, Tobler et al. 1968. 7. Dacey 1964, 1965a, 1965b. 8. Flowerdew 1973a, 1973b. 9. Betak 1974a, 1974b.



for dynamic spatial models, must depend on increased contributions from this line of development.

#### IMPACT OF COMPUTER DEVELOPMENT ON GEOGRAPHIC INFORMATION SYSTEMS

The above discussion of lines of development were based on present-day characteristics, capabilities and costs of computers. The development of computer capabilities, however, continues apace and it is possible to speculate briefly on their impact on geographic information systems.

Without further justification, it will be assumed that computer storage capacities will increase; access times will decrease; the cost per bit of information processed will decrease; higher-order (more human-oriented) computer languages will be developed that allow easy use of computers by people who are not programmers; interactive peripheral devices and supporting software will proliferate; the capacity of low-cost mini-computers will increase; communication networks between computers will increase; the world-wide availability of computers will increase; and the number of people able to use computers will increase. There is sufficient evidence in readily observable trends to be sure that these developments will occur. Others<sup>1</sup> can be foreseen, but are not required for the predictions listed below.

The following developments related to geographic information systems can be expected:

The trend toward creation of large files of non-graphic, spatially coded data will continue and probably accelerate. The 1970-71 Census of U. S. A. and Canada is available to the public in the form of machine-readable summaries of spatially coded small areas. Every town and city in France with a population of more than 20,000 is scheduled to have an urban data system of the line-segment (DIMF) type by 1975. The Ordnance Survey of Great Britain has adopted non-graphic storage for its master map sheets of topographic survey. All hydrological data gathered by the Environmental Protection Agency in U. S. A. (currently 10 years of observations, using  $700 \times 10^6$  bytes of storage) is now in non-graphic form. Many further examples could be provided.

1. Somewhat more esoteric developments might include major changes in computer architecture such as the development of large-scale parallel computers (such as ILLIAC IV), special-purpose computer languages (e.g., for spatial analysis), a wide range of graphic data input/output devices, optical memories, holographic memories, and the development of the ability to process non-coded information.

The use of mini-computers and interactive displays for handling spatial data will rapidly increase, both for teaching techniques and for operational geographic information systems. For example, it is quite practical, even now (1974), to store maps on magnetic tape in inexpensive (\$3.00), standard (Phillips) cassettes, for use with mini-computers.

Improved communication networks between major computers will allow compatible data sources to be correlated.

Picture processing will become economical and will greatly add to the store of spatially coded data on attributes of the earth's surface in non-graphic form.

Multidimensional file structures will be designed, which may encourage interdisciplinary gathering of biospheric data and the development of more complex, dynamic models.

Further speculative comments could be added with little difficulty. However, even with this modest list it can be suggested that the environmental, economic, and social needs of society will encourage use of the large files of non-graphic data being created, and that geographers will take advantage of the related developments in the technology of data handling for the design of geographic information systems.

#### DEVELOPMENT OF GEOGRAPHIC INFORMATION SYSTEMS IN CONTEXT OF THE OVERALL VALUE OF STORES OF SPATIAL INFORMATION

It was noted above that the value of geographic information systems fits within the concept of the usefulness of stores of spatial information. It was suggested that the perceived need for data and their relevance were important elements in the wider concept of usefulness. It would perhaps seem reasonable first to perceive that data are required, then to decide upon the relevant attributes and formats, and thereafter to design and implement the techniques to handle them. In this pragmatic world, however, it is not uncommon for this logical sequence to be disturbed. Techniques for handling data have probably been devised that still await the provision of "relevant" data, and they may wait still longer before the value of those data is perceived. It is equally probable that data are gathered without any

existing techniques for their handling, or any perceived need for them. The comments in the above section of this chapter concerning the future development of geographic information systems have focussed more on the narrow concept of their value. The wider concept and some of the complexities in the relationships between geographic information systems, formal spatial analysis, and decision making in government will be covered in the following part of this chapter.

The conceptual diagram originally presented in Chapter 7 (Fig. 7.1) is useful in illustrating some of these complexities (see Fig. 8.1 below).

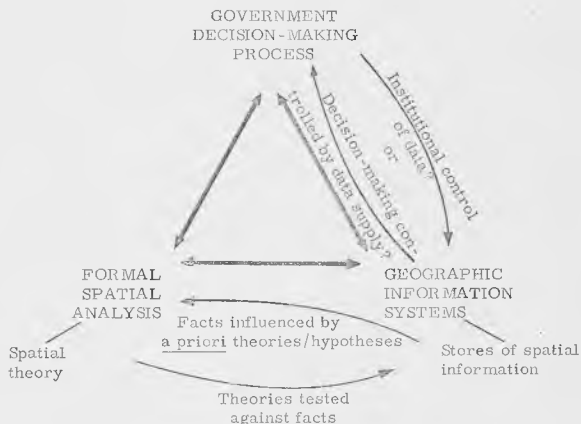


Fig. 8.1. Complexities in the relationships between decision making, geographic information systems, and spatial analysis.

First let us consider the general relationship between stores of spatial information and spatial theory. Burton<sup>1</sup> put the issue clearly in saying that: "The core of scientific method is the organization of facts into theories, and the testing and refinement of theory by its application to the prediction of unknown facts... Theory provides the sieve through which myriads of facts are sorted, without it the facts remain a meaningless jungle. Theory

1. Burton 1963, p. 156.

provides a measure against which exceptional and universal events are recognized. In a world without theory there are no exceptions; everything is unique." Burton<sup>1</sup> also pointed out that "the moment a geographer begins to describe an area, however, he becomes selective (for it is not possible to describe everything) and in the very act of selection demonstrates a conscious or unconscious theory or hypothesis concerning what is significant." Similarly, the gatherers of data and providers of stores of information are selective (for it is not possible to describe everything) and in the very act of data selection use a conscious or subconscious theory or hypothesis concerning what is significant and "relevant". In designing information systems, too, the designers choose between creating different file structures and manipulation facilities on the basis of their conscious or subconscious theory or hypothesis concerning what is significant.

It has already been mentioned that there has been little contact between those concerned with formal spatial analysis and those concerned with the development of geographic information systems. In numerous cases, as has been shown, the systems have been developed by government data gathering and administrative agencies that rely extensively on mental models. This creates no difficulties per se, but the relationship between such data and formal spatial analysis is no longer necessarily the relatively straightforward one of large stores of facts being organized into theory, as the data are only available in prefabricated bundles. In addition, if the stores of spatial information are assumed to represent the total data available and to contain those facts against which theories must also be tested, there is the ever-present possibility of circular argument unless the total data set is suitably partitioned. Less obvious is the danger that each side in such a partition has been influenced by the conscious or subconscious a priori theories or hypotheses of data gatherers and designers of geographic information systems.

The situation is further complicated by the overall relationship between stores of information and decision making in government. Following the above argument, it could be suggested that governmental decision making, particularly at management and policy levels (to the extent that they rely on government sources of data), is influenced by the supply of data that resides in the stores

1. Burton 1963, p. 156.

of information. It can be argued, on the other hand, that the supply of data used in the decision-making process is considerably influenced by the aims and objectives of the agencies of government responsible for decisions. Despite a traditional separation of data gathering agencies (such as census and survey) from line government departments, many data are gathered, and are generally under the control of (and often confidential to), the latter type of agency.<sup>1</sup> Such data are frequently incompatible in time, unit, and spatial character from one institution to another. The selection of contents of many stores of information could, in this sense, be considered to be controlled by such institutions. The overall supply of data is thus partitioned into subsets which are biased by the theories and hypotheses that underlie departmental actions. Policy and programs tend to be explained in terms of such data subsets, and eventual legislation tends to be implemented within the individual jurisdictions of the same institutional framework. There are checks and balances in the system, as have been noted (for example, the policy review agencies and the process of parliamentary government itself), but, to quote Fox<sup>1</sup> in a recent Canadian publication of the (federal) Department of Supply and Services, Bureau of Management, "The procedures of representative government can easily become undemocratic, with Members of Parliament and the general public the victims of the goals and biases of the specialists (and institutions) who control the generation of information on which decisions must be based." Clearly, the relationship between available data, spatial analysis, and government decision is not straightforward.

#### FUTURE DEVELOPMENTS WITHIN THE DISCIPLINE OF GEOGRAPHY

Some of the complexities inherent in the overall relationships have been briefly mentioned. This leads to a re-examination of the linkage between formal spatial analysis and geographic information systems, and consideration of their future interaction. These concerns might be considered to fall within the discipline of geography itself, as illustrated in the next conceptual diagram (Fig. 8.2). One aspect<sup>2</sup> of this broad field of study, the

1. Fox 1973. 2. Other typical examples are: the interaction between geographic information systems and quantitative techniques needed to measure or derive measures for environmental intangibles (for a review of the latter techniques see Coomber and Biswas 1972), and the interaction of further theoretical analysis of the temporal dimension *per se* and the resultant restructuring of space-time models leading to a redesign of geographic information systems (see, for example, the discussion of temporal dimensions by Isard 1970).

better use of a priori knowledge and the attribute-space-time interactions among data, will be used as an example.

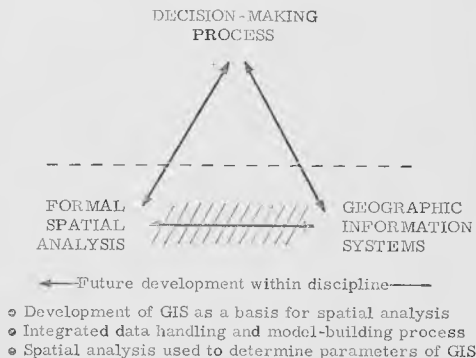


Fig. 8.2. Future developments within the discipline of geography.

While it may be obscure (as pointed out above), it is really not surprising that a priori knowledge influences the contents of stores of information. Although it biases available stores, the approach is necessary and understandable. In fact, in discussing the relationships between formal spatial analysis and stores of information (and in particular the contents and capabilities of geographic information systems), it is of central concern.

James Tobin,<sup>1</sup> quoted by Koopmans, puts the underlying issue clearly by saying that quantitative methods "and developments in high-speed calculating machines make it possible to work with (economic) models of much greater complexity, scope and detail than was formerly thought possible. What is wrong is the poor quality of the pieces that we put together in such models." It was argued in the previous chapter<sup>2</sup> that the contribution made by geographic information systems was to increase the amount and variety of existing spatial data, and increase the capability for manipulating those data. The question still to be answered is whether these developments can affect the "quality of the pieces" that are put together.

1. Tobin 1953. 2. Chapter 7, p. 252.

Koopmans,<sup>1</sup> in his examination of two economic models, noted that "they had not shown the existence of a definite pay-off to further disaggregation... they did not show an increase in flexibility commensurate with their gradual increase in size and in detail. This was particularly true for the input-output model, in which a simple theoretical scheme was impressed on a larger and larger body of defenseless data."

Arguments similar to these could be presented concerning formal spatial analysis. Despite the current activity in model building and testing, there seems to be an undercurrent of dissatisfaction with their outcome, and certainly models are not widely applied to decision making. Although computing capabilities and the potential contribution of geographic information systems have definitely made it possible to incorporate more information in models, the perception of how to take advantage of these capabilities is lacking. It is clear that a mere increase in the amount of data used does not guarantee increased understanding of spatial distribution and interactions.

Koopmans<sup>1</sup> suggests that better use must be made of a priori knowledge, but in a key phrase points out that "such a priori knowledge that we have applies to individual situations". He argues that a high degree of data disaggregation is needed to employ such knowledge, and at the same time a greater flexibility in assumptions is necessary to make the level of disaggregation meaningful. He observes that the degree of disaggregation in models may not be sufficient to permit the use of a great deal of available information and that, in most circumstances, estimates of behaviour parameters obtained from aggregate data do not have the direct and exclusive relationship to the desired individual behaviour parameters. "Thus, the degree of aggregation used does not permit us to exploit the simplest types of a priori information we possess."

These latter thoughts have been more recently confirmed by workers concerned with spatial models, and have been particularly evident in problems concerning human behaviour. Isard and Reiner<sup>2</sup> pointed out that "models of human behaviour over space have been almost entirely oriented to mass probabilistic behaviour." On the basis of extensive experience with migration and diffusion models, Hagerstrand<sup>3</sup> noted that "in spite of the intuitive feeling among all workers in the field (of migration) that micro-environmental factors are

1. Koopmans 1957. 2. Isard and Reiner 1966. 3. Hagerstrand 1970.

important in the decision to move, nearly all models involve only the extrapolation of current aggregate behaviour."

Until now, however, it has generally been assumed that microdata have the inherent problem of being created by activity within the daily range of the majority of the population, and hence are subject to a variability that is avoided in the homogeneity of macrodata. Nevertheless, linkages must be developed between such microdata and the aggregate data if we are to gain an understanding of the large socio-environmental mechanisms that aggregate data fail to give us. One avenue of approach to this problem, suggested by Hagerstrand,<sup>1</sup> is the physical approach involving the study of how events occur in the time-space framework. This implies the availability of geographic information systems whose file structures and languages are considerably further developed than those described in this thesis; it emphasizes the early stage of development of existing techniques.

From these thoughts, however, it is possible to summarize some possible future lines of development that involve interactions between formal spatial analysis and geographic information systems.

- 1) Better use of the a priori knowledge of individual situations involves the use of disaggregated data sets, and implies the theoretical development of linkages between microdata and macrodata, so that understanding of the larger socio-environmental mechanisms can be gained.
- 2) These developments in turn rest on advances in geographic information systems, both to make better use of available stores of information and to handle the attribute-space-time interactions among data. This in turn will depend on substantial improvements in the file structures of geographic information systems. As demonstrated in Chapters 4 and 5, the current file structures are at best straightforward and are the product of system designers' capabilities rather than user demands.
- 3) The development of more efficient file structures will depend to a considerable extent on the theoretical development of multidimensional languages.

This type of mutual development of formal spatial analysis and geographic information systems will have considerable impact on the economic

1. Hagerstrand 1970.



availability of additional relevant information for the understanding of spatial distribution and interactions. The specific approaches within these interacting lines of development must necessarily rely on future work. Theoretical development along these lines, however, assumes access to data. Such access is possible in countries with traditions of open central registers of microdata, such as Sweden. People working in other countries might have to carry out research within the aegis of institutions that gather such data, though it can be noted that geographical information systems to date have largely been built by such agencies. The constraints on wider use of data imposed by social and institutional confidentiality was discussed earlier<sup>1</sup> in the thesis and can present formidable obstacles to investigation.

Somewhat more pragmatic but considerably more complex are the linkages between governmental decision making and the techniques of spatial data handling and analysis. The area of concern is illustrated in the conceptual diagram (Fig. 8.3) below. Some of the possible trends in decision making in response to developments in geographic information systems and spatial analysis are shown, and will be discussed in the text that follows.

#### FUTURE DEVELOPMENTS RELATED TO DECISION MAKING IN GOVERNMENT

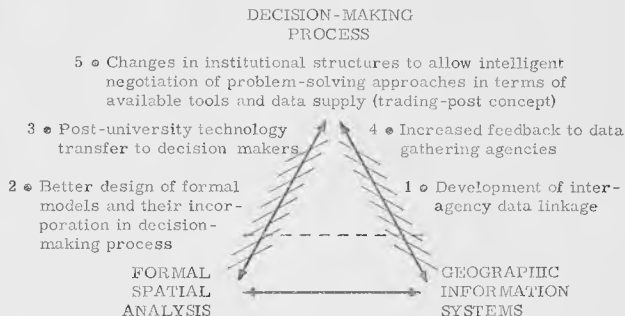


Fig. 8.3. Future developments of decision making in response to developments in geographic information systems and spatial analysis,

1. Ch. 2, pp. 55-56.

Two interrelated developments affecting decision-making institutions underly the need to improve decision making in government. First, as the interdependence of the people of a nation increases, authority and responsibility for governmental affairs appear to gravitate upward from local to regional to national government.<sup>1</sup> This trend is strong, though not without pressures for reversal. Increasing levels of education, population, communication, and standard of living, coupled with the apparently increasing "remoteness" of government, have fostered demands for greater citizen participation in the decisions being made by government (as initially distinct from the overthrow of government). In some cases this pressure, in concert with others, has led toward movements for devolution. Nevertheless, the impact of government decisions on private lives increases.

Secondly, the decisions that face governments are rapidly becoming more complex, as clearly demonstrated by the current problems of regional economic expansion, environmental pollution, landscape degradation, and species preservation, to mention only those most directly related to spatial data. The result of these developments has been a rapid escalation of the need for better decision-making processes, in particular in national government. The simplistic diagram of levels of decision making within government given in Chapter 7 (Fig. 7.4) obviously does not show the strains and discontinuities in the relationships which are inherent in the more complex decision-making tasks. To illustrate these, the diagram on the following page is useful as a starting point for discussion.

It can easily be understood that problems such as environmental pollution may cut across the traditional lines of jurisdiction of numerous government agencies. Given a rational state of affairs, a data base would be available from which data could be extracted, manipulated, and reformatted by handling techniques of various types; the reworked data would be input to selected models, mental or formal, which would apply to the complex problems concerned; and decisions based on the models would be made for subsequent action. Some of the assumed relationships have already been

1. Similarly (but outside the scope of this discussion), this interdependence increases the authority and responsibility of international and global governments, as is witnessed by the creation of entities such as the Common Market in Europe and the growth of multinational corporations. See Fox 1973 for a discussion of the conflict between this trend and the social and philosophical concept of democracy.

examined in this chapter and the reader is generally aware of some of the discontinuities in the progression shown in Fig. 8.4. There is not one data base, but many, frequently incompatible with each other and of various levels of reliability. As has become apparent in Chapters 3 to 7, our capability for handling spatial data in particular is in its very early stages of development.

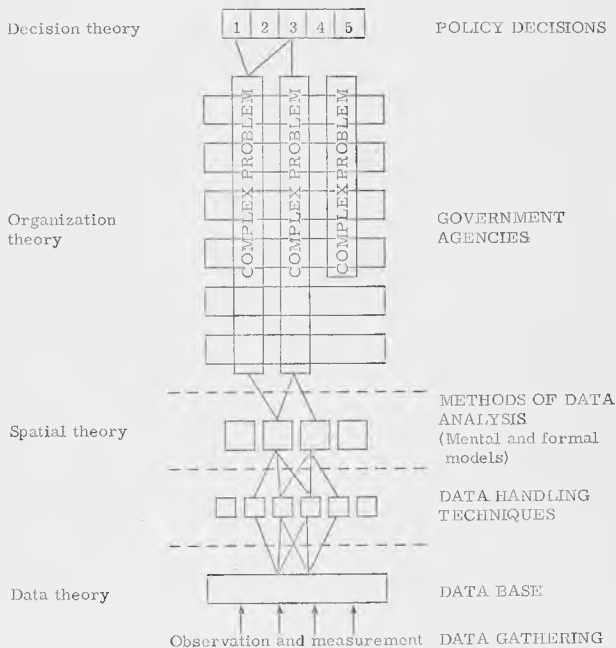


Fig. 8.4. Conceptual sequence of events between data gathering and policy decision making within government.

Similarly, models, both formal and mental, for handling extremely complex problems are in an embryonic stage. With our present institutional

structures, problems tend to be handled from the point of view of individual government agencies, rather than in a concerted manner. Instead of one answer, one might expect as many approaches to a complex problem as there are government departments. The traditional response to this recurring situation is to reorganize the institutions, creating new departments. To deal with critical problems of housing and urban development, one creates a Department of Housing and Urban Development,<sup>1</sup> for problems of regional economic expansion, a Department of Regional Economic Expansion,<sup>2</sup> and for grievous problems concerning the environment, a Department of the Environment.<sup>3</sup>

The arguments in favour of this process are that new jurisdictional boundaries are necessary to handle the new programs, that the problem warrants the full-time concern of a senior minister and staff, or that rationalized fiscal programs can be devised within the new structure. It is suggested that there is another underlying reason: the power of a single institution is necessary to acquire a consistent body of data to apply to the complex problem. Because of the rapidity with which complex problems seem to be emerging, and the inevitable operational inefficiencies inherent in major restructuring of any large organization, governments are attempting to develop techniques that will allow a more efficient response to complex problems, particularly where the provision and handling of data are concerned. One of the difficulties in such development, however, is the lack of adequate theory as a guide to future action. The very categories of theory: data theory, spatial theory,<sup>4</sup> organizational theory, and decision theory, that are central to the government process, are among the least understood and least amenable to rigorous explanation in 1974. Government in this complex age is, even more than before, a grand experiment. Without going further into the overall philosophy of government, it can be observed that at least three major philosophies of government, each covering approximately one-quarter of the world's population (Western world, Marxist-Leninist, Maoist), exist side by side. No one of them is demonstrably "better" than the other in terms of governmental organization. It is suggested that each faces the same weakness in the theory that relates to their process and structure.

1. For example, HUD, U. S. A. 2. For example, DREE, Canada. 3. For example, Department of the Environment in several countries. 4. The restriction to spatial theory here is with respect to the terms of this thesis.

The relationship among the theories underlying the process of government is outside the scope of this discussion, but must be incorporated in future work relating to the development of information systems. In the interim, lines of development that might benefit and have a pragmatic influence on the response of decision making to developments in geographic information systems and spatial analysis will be suggested.

#### Institutional Data Linkages

Perhaps the first line of development that can be suggested is an increase of interagency data linkages. Certainly this goes counter to the concerns of existing institutional power (but if it were pointed out that an alternative was a decline in individual institutional responsibility and the growth of increasingly centralized power in superdepartments, the individual institutions might find such joint action attractive). However, steps<sup>1</sup> toward compatibility of data between agencies are being taken (even if they are being watched extremely warily by the institutions concerned). In fact, the need for institutions to be effective as well as powerful has encouraged the trend toward compatible data.

One approach to this line of development is to create "mini-institutions" which are responsible for gathering interagency data. One of the first examples of this was the Central Statistical Office in Great Britain, during World War II. It was established to resolve arguments arising from conflicting figures given to the Cabinet by various departments, and to provide the framework of national accounting needed to manage the overall demand in the economy according to Keynesian principles. (Note that the stress of war, the need to understand a complex situation, and the specified need for rigorous explanation were factors that induced institutional compliance.) The Business Statistics Office<sup>2</sup> proposed in Great Britain, intended to gather all data from all industrial establishments for subsequent use by all departments needing them, follows the concept of the first example.

A second approach is the establishment of new technical guidelines for data handling. A time of change, such as during the transition from manual to

1. The process has received some impetus from the advent of computers as information processors. As the changeover from manual to computer processing is made, there can be an attempt (sometimes at the technical level rather than the policy level) to make data sets compatible (given guidelines, competence, and resources). Just as easily, however, such a movement can be frustrated, and the incompatibility of data can be enshrined in computer storage. 2. Bray 1970.

computer data processing, provides opportunities for such innovations. Without legislative underpinnings, however, this approach is less likely to be successful in a time of methodological stability. An example is the current effort of the Department of Environment in Great Britain to guide the spatial encoding of urban data toward point data systems.<sup>1</sup> A third approach is the provision of technical expertise, the central training of technicians responsible for designing new data structures. Again, this approach is primarily useful in a time of methodological change. It is being attempted in a program extended to 225 cities in the U. S. A.,<sup>2</sup> and underlies the program of the Centre d'Études et de Recherches sur les Collectivités Locales en Europe (C. E. R. C. L. E.)<sup>3</sup> proposed in France for the training of middle levels of government staff in the technology of data handling. A third approach is that of the "pilot study". An example of this can be taken from France: the Opération Pilote Interministérielle sur les Données Administratives<sup>4</sup> has been established at national level in departments of government with the specific objective of integrating institutional data for planning in the South of France. The aim of this exercise is eventually to make the data in the various government agencies compatible and not to establish a new agency. The fourth approach is to require compatible data sources by legislation, as a prior condition for receiving funds from a central source. An example of this is the Jackson Bill<sup>5</sup> currently being debated by the U. S. Senate, which will require every state to have a survey of present land use, computer-stored in a compatible format, as a basis for future grants for federal land planning and development.

The degree to which these approaches are needed and will achieve success varies. As pointed out in Chapter 2, the existing level of data compatibility between government departments varies markedly from one country to another. The 400-year-old tradition of open Central Registers in Sweden contrasts with the more confidential and flexible approach to data gathering in Great Britain, which contrasts in turn with the immense complexities due to size and a federal system of governments in the U. S. A. Nevertheless, the development of inter-institution data linkages will have considerable impact on the supply of data available and on the type of decision making

1. Department of Environment, U.K. 1973. 2. U. S. Bureau of Census 1973.  
 3. The author is a member of the steering committee for this program.  
 4. Salmons 1974. 5. National Land Use Policy and Planning Assistance Act, S268, 93rd Congress of U. S. A., 1st Session. See also Calkins et al. 1973.

that can benefit from improved techniques for handling spatial data and improved formal spatial analysis (and vice versa).

#### Design of Formal Spatial Models

The development of types of formal spatial analysis that are more valuable to government decision making than those currently available will certainly be encouraged by developments within the discipline of geography and by comparable increases in supplies of compatible data within government. Real improvement, however, must rely to a great extent on increased communication between model designers and decision makers. It is difficult to take an objective view of the situation and decide which group is hesitant, and why. Simple models, which relate to the intuitive models of decision makers, frequently seem to be inadequate for decision making because of the inflexibility of their inherent assumptions (assumptions used very flexibly in mental models). More complex models, which attempt to be flexible with respect to the same assumptions, do not relate to the intuitive mental models of decision makers, are incomprehensible to an untrained mind, and, despite their state of development, are still awkward tools. The relationship is complicated by the fact that governmental decisions are frequently made in urgent response to problems (they operate in "real time") while the design and testing of relevant models may take more time than is available before such decisions have to be made (they operate "off-line"). There are some examples of models being integrated with the decision-making process within an institutional framework. The urban transportation models based on the work of Garin and Lowry are well established in Pittsburgh, U. S. A. Sub-regional growth and allocation models are in routine use in England in Leicestershire, Cheshire, and Hampshire. The Swedish Government employs quite sophisticated models of regional growth as an integral part of policy and planning decisions. The trend is, however, in its infancy. The current overall situation could probably be summarized by saying that there are few formal models useful to the decision-making process on a recurring basis, and there is a widespread lack of comprehension of spatial analysis on the part of decision makers.

It can, however, be noted that the development of spatial models in recent

years has closely followed the concerns of society. The widespread development of transportation and urban models has been in response to governmental problems of transportation and urban development. Doubtless the number of environmental models will increase in the near future. In the opinion of the author, an expressed need for explanation and the funds to support research elicits a strong response from model designers.

Assuming that capabilities for data gathering and handling have been developed, three factors are needed to encourage the development of better models for decision making: increased perception, anticipation, and comprehension on the part of decision makers. It is clear that a decision maker must be able to perceive that formal analysis can apply to the tasks that face him. Before such techniques can be devised, the need for them must be anticipated, to allow the lead time required for model development. This involves a change in mode of government from one of response to crises to their anticipation, which in itself will be a major change for governments. The third factor, comprehension, implies that the formal training of decision makers includes subjects that allow him to understand (and depend upon) the techniques being used.

Of these three factors, anticipation is probably the least critical, as to some extent it can be a cooperative process between model builders and decision makers without close contact. The model builders themselves can, to some extent, observe society's trends and direct research accordingly. The factors of perceived need for formal analysis and its subsequent comprehension, however, rely mainly on the training of the decision makers.

#### Post-University Transfer of Technology to Decision Makers

One can hope that current courses of higher education will provide an adequate supply of well-trained people who will eventually participate in the decision-making levels of government. Of more immediate concern are the people who have completed their formal academic training and who will influence decision making in government in the next decade. The experience of the U. S. Bureau of the Census in a recent series of Workshops for training local government officials in the techniques of the DIME<sup>1</sup> system illustrates the situation. The objectives of the workshops (each 5 days long)

1. See Ch. 4, pp. 116-122.



were to ensure "that the inherent capabilities of the DIME system facilities can be utilized by those implementing it in working environments and... to provide a feedback process so that the continuing design of the DIME system facilities can be carried out in response to expressed user needs and problems."<sup>1</sup> The workshops were held at regional centres in the U. S. A. (Indianapolis, Atlanta, Pittsburgh, Boston, and San Francisco). Approximately 200 people from middle levels of management in urban and regional government attended. The author was a member of the advisory committee for the workshops and attended four of them. Quite simply, the workshops showed that a key factor in use of the DIME system was the ability of the users to approach planning problems in a manner that required the type of data the system could provide. As the DIME system is fairly well developed, quite flexible, and reasonably efficient, the limiting factor was the competence of the planners to approach their problems in quantitative terms, to think in spatial terms, and, in general, to be able to take advantage of the data system capabilities. This lack of training has equal impact on the perceived need for and understanding of formal spatial analysis, and the design and use of geographic information systems to provide data for both formal and mental models.

The fact remains that few structures in our society allow the post-university transfer of advanced technology to administrators and decision makers, who most need to be aware of its potentiality. It is clear that one line of development must be an increase in formal and informal post-university education, whether it be in the form of sabbaticals, summer courses, short courses, seminars, workshops, audio-visual presentations, computer-aided instruction, programmed instruction, self-teaching manuals or an appropriate mixture. Again, the detailed approach to be followed must depend on future study. What is important is that such training becomes more readily available, and becomes an integral part of the career path of decision makers. As an interim measure, it is suggested that those concerned with formal spatial analysis, and in particular with the design of geographic information systems, understand that it is necessary to build user competence in quantitative methods of problem solving, as well as to build information systems and formal analytical capabilities.

1. See U. S. Bureau of Census 1973.

### Interaction Between Data Gatherers and Data Users

The mandate of departments of government that traditionally gather data (census and survey) is to provide data relating to the perceived needs of government impartially to all departments of government (and to a lesser degree, to all other public and private users). This is a very broad mandate. As the tasks facing such data gathering departments have increased, and as the delay between the need for data and their receipt has become more critical, the line departments of government have had increasing reason to use their own administrative networks (frequently more extensive than those of data gathering agencies) to collect data related to their own operations. Nevertheless, data gathering agencies gather large amounts of data. Although such data necessarily are biased by the agencies' interpretation of the "perceived needs of government", they are otherwise impartial and internally consistent.

There are two substantial drawbacks to the present situation. First, the departments' interpretation of and reaction to the perceived needs of government may cause data to be gathered without later use; and once data collection has begun, there is a tendency for it to continue if only for the sake of continuity.<sup>1</sup> Secondly, the process of feedback from government, the specification of the perceived needs of government by government, is substantially filtered through the intervening line departments or does not occur at all. These drawbacks place a great deal of responsibility on the shoulders of the data gathering departments and perhaps increase the danger of irrelevant data being gathered or of line departments being encouraged to gather their own "relevant" data. At the same time, the techniques and capabilities for gathering data are rapidly improving. In the survey profession in particular, there have been spectacular developments in recent years in the capability for gathering topographic and thematic spatial data. The advances in conventional aerial photography, automatic photogrammetry, ortho-photography, multispectral sensing, satellite sensor platforms, helicopter reconnaissance surveys, spatial sampling techniques, and automatic cartography represent a great increase in the technology of data gathering and display, which will be exercised primarily by traditional

1. The inertia in the system is not necessarily bad. Continuity of data sets is a valuable property. The inertia, however, affects both useful and useless data. The problem lies in establishing criteria for distinguishing between them.

data gathering institutions: We are undoubtedly in a position where we can gather huge amounts of data, which we may or may not be able to handle, and which we may or may not need. Clearly, increased feedback from the policy making levels of government is needed to identify which data could profitably be collected and, perhaps equally importantly, to provide the criteria for delimitation.

Such revision will not occur quickly. No mechanisms are in place to improve the feedback, and there is substantial momentum to the gathering of current data sets. Increases in the total supply (of both wanted and unwanted data) and possible overloading of storage and handling capacity for some types of data, are likely to occur before better feedback has a greater influence in the selection of data to be gathered.

Perhaps the first improvements in feedback to the data gathering agencies that provide spatially coded data will come from the use of geographic information systems. Only when data are actually read<sup>1</sup> and used in mental or formal models, and it can be shown that content or format changes are necessary or that other data might be profitably collected with the same expenditure of funds, is there any likelihood that persuasion can be brought to bear on traditional data gathering agencies to amend their standard forms of operation.

The second level of improvement in feedback will come when improved spatial models are integrated into the decision-making process and can be used to specify data requirements. This implies a change in the decision-making process as we know it today. A substantial degree of integration of formal techniques in the decision-making process would be needed before their data needs would represent a strong argument for changing established data gathering practice. The possibility of such changes in the decision-making process is examined below.

#### Interaction Between Data Specialists and Decision Makers

Several strands of thought run through the discussion in the latter part of this chapter. The decision-making process, as has been seen, relies primarily on mental models. Formal models have a valuable part to play in decision making, but decision makers generally lack a clear comprehension

1. The benefit of even the current level of techniques for handling spatial data is that it is now possible to examine the data previously collected, but not before amenable to economic use.

of their capabilities. Geographic information systems have been developed mainly in response to sources of data and the needs of mental models. They will be further developed to handle increases in the gathering of spatial data, in the stores of spatially coded, machine-readable, existing data, and in the demands of formal models. The concerns of government are observably becoming more complex and improvements in decision-making processes are needed within the organizational structure of government.

In the short term, there is an obvious need to increase the communication between decision makers and those concerned with the provision, handling, and analysis of data. Communication in this sense does not simply refer to an exchange of information, but implies "negotiation". The concept underlying this suggestion is that decisions are not made from the data themselves or from data provided by mental or formal models, but from the images that the data create. In the absence of data, or any form of analysis of data, the image may be blank or preconceived, but that does not alter the fact that the decision is made from the image.<sup>1</sup> The impact of data, and the analysis of data, on a decision is thus the extent to which the image is created or changed by the data or analysis. The function of the designer of an information system or a formal model in decision-making terms is to provide output that has impact on the image. The flexibility of the process (in the content and manipulation facilities of information systems, and in the types of formal analysis) is that images may be generated or changed by one of any number of subsets of the data or types of analysis needed to quantify the image completely. The required communication process is thus one that negotiates the interface between the image and the data; that negotiates between decision making and the processes of information system design and formal analytical design. The outcome of negotiating the interface between decision and data would be a new view and understanding of data that those creating and maintaining data banks and information systems frequently lack. Similarly, it would measurably assist the design of formal models that could be integrated into the decision-making processes.

In general, the structures that would allow the needed communication between

1. If a person knows absolutely nothing about a subject, e.g., life in outer space, he is quite capable of making a decision about it, either from a blank image or an image based on absolutely no evidence. Given data, he can process them either mentally or formally and arrive at a conclusion, which may or may not alter his image. In this sense an image is an abstraction of a mental model; the former is a mental state, the latter a mental data processing procedure.

data specialists and decision makers are not now common in government at local, regional, or national levels. The development of such structures is a real need of local government. The need is critical at regional and national levels particularly, to permit interaction between those concerned with multi-jurisdictional policy and those concerned with provision and analysis of data. Such a structure could include the function of a "go-between": essentially, someone who is trusted by the decision maker but who understands the technology of information systems and formal data analysis. The function of an "interpreter" of the technology is also needed; one who could also augment such technology with supplementary forms of analysis or supply of data, but whose basic requirement would be the sensitivity to understand the difficulties of communication in dealing with decision makers, especially when they lack sophisticated technical training. The concept is one of an "intellectual trading post" for transactions between need and capability.

There is some evidence that such structures are in embryo form in some countries, at some levels of government. In complex bureaucratic organizations, the trading-post concept can be implemented invisibly within any, or several, existing frameworks. In less complex governments, the structures would be more evident, particularly near the policy level of government where they are most needed. The author was part of such a unit in the Secretariat to the Cabinet of the Government of Nova Scotia between 1969 and 1971. This was an entirely new formal institution within a provincial government in Canada. At the national level in Canada, it can be observed that the staff in the Prime Minister's office has increased substantially in recent years (from 20 to over 100 persons).

The concept of the trading post is only one of several methods for creating a structure that would allow for effective communication. Others can be thought of without much effort. Regardless of the actual structure, however, the need for increased communication is apparent. Such communication will not occur without effort; the effort should, however, be a temporary phenomenon that will decrease as the decision-making processes and agencies responsible for provision and analysis of data adjust to new demands, and as training changes the skills of the participants.

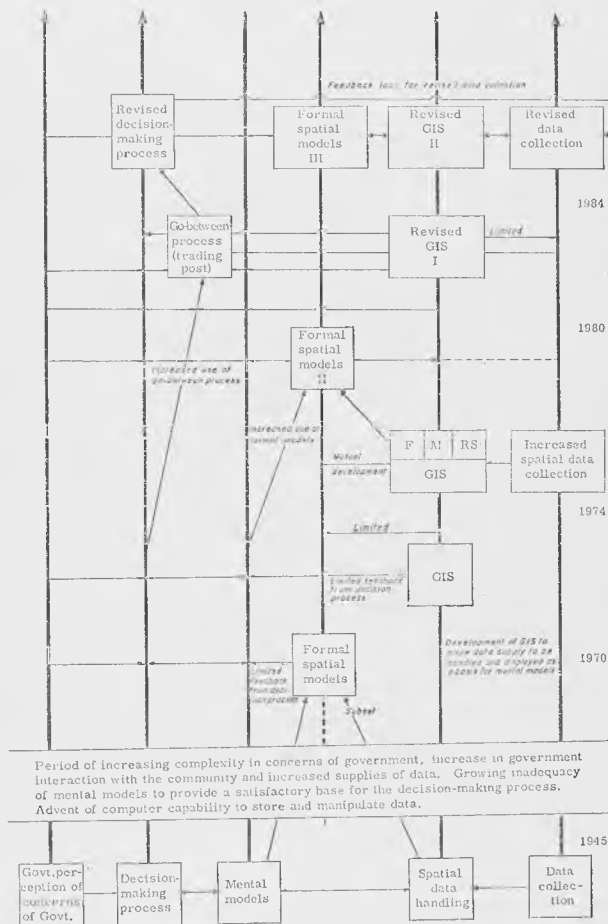


Fig. 8.5. Interactions between data collection, spatial data handling, mental and formal spatial analysis, and decision making, over time.

The linkages between the various strands of thought noted above are complex and vary over time. They have been summarized in Fig. 8.5. For the purposes of discussion, the increases in communication needed are represented by a trading-post structure.

The diagram is conceptual and concentrates on the central concerns of the thesis: those of spatial data handling, spatial data analysis, and decision making in government. Dates are indicated on the right side of the diagram. The time scale is approximately linear between 1970 and 1984, but not rigorously so. Before 1945, the data collection process provided data for spatial data handling, which was exclusively manual at the time. This, in turn, provided input to (and received direction from) the exclusively mental models, which had input to (and received direction from) the decision-making process, which operated in response to the government perception of the concerns of government.

For all the processes, the lines of development continue (vertically) from before 1945 to the top of the diagram. These processes are all thus still with us and are likely to continue in the foreseeable future. If a process in the diagram undergoes significant change, a new box has been inserted on the process line to indicate the change. Otherwise, the diagram is mainly concerned with changes in the interactions between the processes.

For purposes of simplicity, the many variables that enter into the process of perception of governmental concerns are not shown. Attention is focussed on the relationships between data collection, handling, and analysis, and decision making.

The period between 1945 and 1970 saw increasing complexity in the concerns of government, increase in governmental interaction with the community, and increased supplies of data. Although mental models were still the primary form of analysis, the potential contribution of formal spatial models to decision making became apparent. The computer became available as a calculating and information processing tool.

Drawing from mental models and subsets of data, derived from existing stores of data largely by manual processes, formal spatial models were

introduced as elements in the decision-making process. They made a limited contribution to mental models and to decision making, and there was limited feedback. In this period, techniques for handling spatial data were slowly taking advantage of computer capabilities and developed into the early forms of geographic information systems (geographic information systems in this sense including all processes of handling spatial data, manual and computer-aided). These made their first contributions to mental models and hence to decision making, and there was limited feedback from the decision-making process. Between 1970 and 1974 they made a limited contribution to formal spatial models. This sequence brings us to the current (1974) state of affairs.

The interactions between processes that follow are speculative, but reflect the themes identified in the thesis.

Increases in the collection of spatial data encourage the development of geographic information systems. Geographic information systems tend to develop on separate lines: some (RS) respond to the picture-processing demands of remote sensing; some (M) continue the traditional function of providing input to mental models; some (F), in response to mutual development of formal spatial analysis and data handling, are primarily used to provide input to formal spatial models. In this last type the first experiments with new languages will be made and the temporal variables will receive greater attention. During this period, communication with decision makers (the go-between process) will increase, and the use of formal spatial models in decision making will slowly become more frequent.

Formal spatial models II represent these tools developed to the point where they are considerably more aligned to the decision-making process. They continue to make input to mental models and encourage the development of the go-between process. They make and receive input from the mainstream of development of geographic information systems, and have a limited influence in the collection of data. Geographic information systems, however, are still making their major contribution to mental models for decision making.



At the level of revised geographic information systems I, the systems will probably have benefitted substantially from development of data theory and new spatial languages. A much greater integration of data sources and formats will be possible. Input will be made either to formal models and thence, by way of mental models, to the go-between process, or directly to mental models and the go-between process. Most importantly, there will be a strong feedback to the design of information systems from the go-between process. The feedback to the data collection agencies will still be limited.

After the next decade, it is assumed that many of the communication aspects of the go-between process will have been incorporated into what will then be standard forms of decision making. This will involve communication with a third generation of spatial models integrated into the decision process, and with geographic information systems similarly aligned. The main result of these developments will be strong feedback to the processes of data collection and a revised and selective process of data collection.

It is emphasized again that these interactions are speculative. Some comments on some of the timings envisaged might, however, be worthwhile. The relatively slow implementation of the go-between process is based on a view of the inherent inertia of government institutions and the likelihood that there will have to be demonstrably better formal models and geographic information systems before the need for increased communications will be substantiated. The prediction that development of decision-oriented geographic information systems will be late (circa 1983) is based on the lack of current research in the underpinnings of theory and language of the process. The development of vastly increased computing power<sup>1</sup> would ameliorate the situation, but would not necessarily provide efficient solutions to the underlying problems. The development of a strong feedback loop to the traditional process of data collection must await, and follow, the solid establishment of the other processes. No effective persuasion can be brought to bear while manipulation processes, formal spatial analysis, and decision processes themselves are rapidly evolving.

1. The development of large-scale parallel processing computers similar to the experimental ILLIAC IV at the University of Illinois, or the capability of developing multiple-cpu interfaces, would allow two-dimensional data to be handled with more facility than is currently possible. There is no evidence, however, that such developments in computers can be anticipated in the near future.

The diagram represents one version of many possible patterns of interaction. Perhaps its main function is to stress that such interaction is essential if rational development of stores of information, geographic information systems, formal spatial models, and decision making is to go forward.

#### SUMMARY

These two final chapters of the thesis have briefly examined the inter-relationships between geographic information systems, formal spatial analysis, and decision making in government.

In Chapter 7, the characteristics of formal spatial analysis and then governmental decision making were discussed on the basis of their needs for spatial data handling. The various categories of geographic information system identified in the previous chapters were considered in terms of the contribution they make to these processes. The varying contributions of the different types of system were made apparent and the institutional basis of their development was demonstrated. The overall value of geographic information systems was examined.

In Chapter 8, lines of future development were postulated, assuming present and future computing capabilities. Then, on the basis that geographic information systems are not simply processing tools but that their usefulness is inherently aligned to their broader function as stores of information, their development was examined within the broader context. Some of the general complexities in the relationships between spatial theory, stores of spatial information and the decision process in government institutions were noted.

The future development of interaction between spatial data and formal spatial analysis within the discipline of geography was postulated. It was seen that easy manipulation of particular sets of microdata provides scope for improvements in the speed and precision with which hypotheses are tested. This facility, more than the volume of data itself, will provide the main contribution of geographic information systems to the discipline of geography, especially when directed toward problems of dynamic analysis and supported by the active development of suitable multidimensional languages. The

future development of decision making in response to developments in geographic information systems and spatial analysis was also considered and some lines of development were postulated. It was made clear that the development of geographic information systems as described in this thesis must be viewed not as an isolated process, but as an integral part of a very large structure of data gathering, data analysis, and decision making. The theoretical underpinnings of the structure, in particular theory applying to data, languages, spatial analysis, organization, and decisions are not strong. It is suggested that rational development of both the theoretical basis and pragmatic application will benefit from exploration and recognition of their interdependence, and from a substantial increase in interaction between the techniques of data gathering, data handling, data analysis, and decision making.

# REFERENCES

- Alexander, C. 1964. Notes on the synthesis of form. Harvard University Press, Cambridge, Mass.
- . 1967. The city as a mechanism for sustaining human contact. Chapter 4 in Ewald, W. R. Jr., Environment for man: The next fifty years. Indiana University Press, Bloomington, Ind.
- , and Chermayeff, S. 1966. Community and privacy. Penguin Books, Harmondsworth, Middlesex.
- Alfredsson, B., Salomonsson, O. and Selander, K. 1970. FRIS A:1. A spatial information system - A pilot study. Introduction. Central Board for Real Estate Data, Sundbyberg, Sweden.
- Almdinger, V. 1973. Urban Decision Systems Inc., Los Angeles, Calif. Personal communication.
- Alonso, W. 1964. Location and land use: Toward a general theory of land rent. Harvard University Press, Cambridge, Mass.
- Amidon, E. L. 1964. A computer oriented system for assembling and displaying land management information. Research paper PSW-17. United States Department of Agriculture, Forest Service, Berkeley, Calif.
- . 1966. An alphanumeric map information assembly and display system for a large computer. Research paper PSW-38. United States Department of Agriculture, Forest Service, Berkeley, Calif.
- . 1972. VIEWIT. Pages 943-946 in Tomlinson, R. F., ed., Geographical data handling. IGU Commission on Geographical Data Sensing and Processing, Ottawa, Ont.
- Amsterdam, R. 1968. GIST: A geographic information system for New York City - Preliminary design. City of New York, Office of the Mayor, Office of Administration, New York, N. Y.
- . 1969. The concept of data network and the development of GIST. City of New York, Office of the Mayor, Office of Administration, New York, N. Y.
- Appel, A., Dankowski, T. P. and Dougherty, R. L. 1968. Aspects of display technology. IBM Systems Journal 7(3 and 4):176-187.
- Arms, S. 1967. Maps and computers: Locational information processing for planning and government. Columbia Region Association of Governments, Metropolitan Planning Commission, Portland, Ore. (Out of print.)
- . 1968. MAP/MODEL system: Technical concepts and program descriptions. Columbia Region Association of Governments, Portland, Ore.
- . 1969. MAP/MODEL system: Data entry and retrieval. Bureau of Governmental Research and Service, University of Oregon, Eugene, Ore.

- Arms, S. 1970. MAP/MODEL system; System description and user's guide. Bureau of Governmental Research and Service, University of Oregon, Eugene, Ore.
- Atelier Parisien d'Urbanisme. 1972. Map of Paris. Pages 939-942 in Tomlinson, R. F., ed., Geographical data handling. IGU Commission on Geographical Data Sensing and Processing, Ottawa, Ont.
- Aumen, W. 1970. A new map form; Numbers. International Yearbook of Cartography 10:80-84.
- Balchin, W. V. G. and Coleman, A. M. 1967. Cartography and computers. The Cartographer 4(2):120-127.
- Barb, C. E. 1968. Street address conversion system: An automated, geocoding system for urban analysis. Papers from the 6th Annual Conference of the Urban and Regional Information Systems Association, Clayton, Miss.
- Barracough, R. E. 1972. National geocoding system. In Geocoding '72. Papers of the Geographic Base File Special Interest Group, 10th Annual Conference, Urban and Regional Information Systems Association, San Francisco, Calif.
- Barrett, J. 1973. Department of Data Systems, Metropolitan Washington Council of Governments, Washington, D. C. 20035. Personal communication.
- Batty, M. 1970. Spatial theory and information systems. Working paper USRU-WP-310. Urban Systems Research Unit, University of Reading, Reading, England.
- Becker, P. C. 1971. The rise and fall of the master parcel file. Proceedings of the 9th Annual Conference of the Urban and Regional Information Systems Association, New Orleans, La.
- Bengtsson, B. -E. and Nordbeck, S. 1964. Construction of isarithms and isarithmic maps by computers. Nordisk Tidskrift for Informations-Behandling 4(2):87-105.
- Berry, B. J. L. 1961. City size distributions and economic development. Economic Development and Cultural Exchange 9(4):573-588.
- . 1967. Geography of market centers and retail distribution. Prentice-Hall, Englewood Cliffs, N. J.
- and Garrison, W. L. 1958. Alternate explanations of urban rank-size relationships. Annals of the Association of American Geographers 48:83-91.
- and Marble, D. F., eds. 1968. Spatial analysis: A reader in statistical geography. Prentice-Hall, Englewood Cliffs, N. J.
- Bertin, J. 1970. La cartographie dans "La civilisation de l'informatique." International Yearbook of Cartography 10:85-94.
- Betak, J. F. 1974a. The conceptual locus of a two-dimensional language: Some implications for human responses to two-dimensional displays. Department of Geography, McMaster University, Hamilton, Ont. Mimeograph.

- Betak, J. F. 1974b. Syntactic and subjective complexity: A simple test of a two-dimensional language model. Department of Geography, McMaster University, Hamilton, Ont. Mimeograph.
- Bickmore, D. P. 1966. Experimental cartography project. Interim report to the Natural Environment Research Council. Royal College of Art, London, England.
- \_\_\_\_\_. 1971. Automatic cartography and planning. Architectural Press, London, England.
- Boehm, B. W. 1967. Tabular representations of multivariate functions - with applications to topographic modeling. Proceedings of the 22nd National Conference, Association of Computing Machinery. pp. 403-415.
- Boehne, R. C. 1967. Digital simulation of environmental limitations of inter-visibility. In Hollingdale, S. H., ed., Digital simulation in operational research. American Elsevier Publishing Co., New York, N. Y.
- Boeing Computer Services, Inc. 1972a. Natural resource information system, volume I: Overall description. Boeing Computer Services, Inc., Seattle, Wash.
- \_\_\_\_\_. 1972b. Natural resource information system: Design analysis. Boeing Computer Services, Inc., Seattle, Wash.
- Bogue, D. J. et al. 1957. Subregional migration in the United States: 1935-40, volume 1. Scripps Foundation Studies in Population Distribution, Oxford, Ohio.
- Borchert, J. R. 1969. Minnesota lakeshore development study. Department of Geography, University of Minnesota, Minneapolis, Minn.
- Boyle, A. R. 1970a. Computer aided compilation. Paper presented at the 9th Annual Hydrographic Conference. Canadian Hydrographic Service, Department of Energy, Mines and Resources, Ottawa, Ont.
- \_\_\_\_\_. 1970b. Automation in hydrographic charting. The Canadian Surveyor 24(5):519-537.
- \_\_\_\_\_. 1971. Computer aided map compilations. National Research Council 2nd Seminar, Report on Man/Machine Communication. National Research Council, Ottawa, Ont.
- \_\_\_\_\_. 1972. A character recognition device for soundings on navigational charts. Surveying and Mapping 32(1):27-35.
- Bray, J. 1970. Decision in government. Victor Gollancz, London, England.
- Broadbent, T. A. 1969. Some techniques for regional economic analysis. CES-WP-23. Centre for Environmental Studies, London, England.
- Brown, T. 1970. Automatic line following. M. Sc. Thesis, University of Saskatchewan, Saskatoon, Sask.
- Bruer, P. M. L. 1967. The Erie project narrative. Paper presented at University of Western Ontario, Department of Geography Conference on the Erie Project. University of Western Ontario, London, Ont.

- Burch, J. G. Jr. and Strater, F. 1974. Information systems: Theory and practice. Hamilton, Santa Barbara, Calif.
- Burgess, E. W. 1925. The growth of the city. In Park, R. E. and Burgess, E. W., eds., The city. Chicago University Press, Chicago, Ill.
- Burton, I. 1963. The quantitative revolution and theoretical geography. The Canadian Geographer 7(4):151-162.
- Calkins, H. W. 1967. Ottawa street address conversion system. Highway Research Record 194:96-102.
- \_\_\_\_\_. 1973. Urban Systems Laboratory, University of Washington, Seattle, Wash. Personal communication.
- \_\_\_\_\_, Robertson, J. L. and Wierman, S. S. G. 1973. The legislative mandate for state land information systems. Papers from the 11th Annual Conference of the Urban and Regional Information Systems Association, Atlantic City, Ga.
- Carnap, R. 1958. Introduction to symbolic language and its applications. New York, N. Y.
- Carrothers, G. A. P. 1956. An historical review of the gravity and potential concepts of human interaction. Journal of the American Institute of Planners 22:94-102.
- Cassetti, F. and Semple, R. K. 1968. A method for stepwise separation of spatial trends. Discussion Paper 2, Michigan Inter-University Community of Mathematical Geographers. Department of Geography, University of Michigan, Ann Arbor, Mich.
- Cater, E. A. 1970. Information needs for planners: A survey. Working paper USRU-WP-4. Urban Systems Research Unit, University of Reading, Reading, England.
- Chapin, F. S. Jr. and Weiss, S. F. 1962a. Factors influencing land development. Institute for Research in Social Sciences, University of North Carolina, Chapel Hill, N. C.
- \_\_\_\_\_, eds. 1962b. Urban growth dynamics. John Wiley & Sons, New York, N. Y.
- Cheshire, A. 1970. Statistics for planning. Working papers USRU-WP-5 and USRU-WP-6. Urban Systems Research Unit, University of Reading, Reading, England.
- Chisholm, M. 1961. Rural settlement and land use. Hutchinson University Library, London, England.
- Chorley, R. J. and Haggett, P. 1965. Trend-surface mapping in geographical research. Transactions of the Institute of British Geographers 37:47-67.
- \_\_\_\_\_, eds. 1967. Models in geography. Methuen & Co., London, England.
- Christaller, W. 1935. Die zentralen Orte in Süddeutschland. G. Fischer Verlag, Jena. The central places of southern Germany. Translated by C. Baskin, 1966. Prentice-Hall, Englewood Cliffs, N. Y.

- City of Vancouver. 1965. 'An information retrieval system for urban areas. City Engineering Department, Vancouver, B.C.
- Clark, R. A. 1970. Computer generated profile maps. Proceedings of the Semiannual Convention of the American Society of Photogrammetry, Denver, Colo.
- Cleland, D. S. and King, W. M. R. 1969. Systems, organizations, analysis, management: A book of readings. McGraw-Hill, New York, N. Y.
- Cliff, A. D. and Ord, J. K. 1973. Spatial autocorrelation. Pion Ltd., London, England.
- Cole, J. P. and King, C. A. M. 1968. Quantitative geography: Techniques and theories in geography. John Wiley & Sons, London, England.
- Comprehensive Planning Organization. 1971. Proposed information system for planning, volume I. Comprehensive Planning Organization Job 0050. Comprehensive Planning Organization, San Diego County, San Diego, Calif.
- Cooke, D. 1973. Urban Data Processing Company, Cambridge, Mass. Personal communication.
- Coomber, N. H. and Biswas, A. K. 1972. Evaluation of environmental intangibles: Review of techniques. Environment Canada, Ecological Systems Branch, Research Coordination Directorate, Policy Planning and Research Service, Ottawa, Ont.
- Coombs, C. H. 1964. A theory of data. John Wiley & Sons, New York, N. Y.
- Coppock, J. T. 1962. Electronic data processing in geographical research. *The Professional Geographer* 14(4):1-4.
- \_\_\_\_\_ and Johnson, J. H. 1962. Measurement in human geography. *Economic Geography* 38(2):130-137.
- Corbett, J. and Farnsworth, G. 1971. Theoretical basis of dual independent map encoding (DIME). Papers from the 9th Annual Conference of the Urban and Regional Information Systems Association, New Orleans, La.
- Coventry City Council. 1970. Point data system. Report, March 1970. Department of Architecture and Planning and City Treasurers Department, Coventry City Council, Coventry, England.
- Cripps, E. L. 1970. A comparative study of information systems for urban and regional planning, 1: Scandinavia. Working paper USRU-WP-7/0. Urban Systems Research Unit, University of Reading, Reading, England.
- Dacey, M. F. 1964. Individuation in substance and space-time languages and Carnap's classification of rules of designation and positional coordinates. Memo, GSI (mfd) 1 and 2. Regional Science Department, University of Pennsylvania, Philadelphia, Pa.
- \_\_\_\_\_. 1965a. The geometry of central place theory. *Geografiska Annaler Series B* 47:111-124.



- Dacey, M.F. 1965b. Some observations on a two-dimensional language. Technical Report 7 of Office of Naval Research Task No. 389-142, Contract NONR 1228(35). Department of Geography, Northwestern University, Evanston, Ill.
- and Marble, D.F. 1965. Some comments on certain technical aspects of geographic information systems. Technical Report 2 of Office of Naval Research Task No. 389-142, Contract NONR 1228(35). Department of Geography, Northwestern University, Evanston, Ill.
- Dangermond, J. 1971. Polygon information overlay system. Comprehensive Planning Organization, San Diego County, San Diego, Calif.
- 1973. Technical alternatives for geographical information systems at state/regional level. Papers from the 11th Annual Conference of the Urban and Regional Information Systems Association, Atlantic City, Ga.
- Dayhoff, M.O. 1963. A computer map program for X-ray crystallography. Communications of the Association for Computing Machinery 6(10): 620-622.
- Delaware County Planning Commission. 1970. Delaware County Planning Commission information system. Delaware County Planning Commission, Media, Pa.
- Department of the Environment, U.K. 1973. Manual on point referencing properties and parcels of land. Techniques for planning. Her Majesty's Stationery Office, London, England.
- Deutsch, K.W. 1966. The nerves of government: Models of political communication and control. New York, N.Y.
- Diello, J. 1970. The evolution of automation in cartography at Rome Air Development Center. Air Force Systems Command, Griffiss Air Force Base, Rome, N.Y.
- , Kirk, K. and Callander, J. 1969. The development of an automated cartographic system. The Cartographic Journal 6(1):9-17.
- Dodd, G.G. 1969. Elements of data management systems. Computing Surveys 1:117-133.
- Dowd, R. 1973. Department of Hospitals, Los Angeles County, Los Angeles, Calif. Personal communication.
- Downs, R.M. and Stea, D., eds. 1973. Image and environment: Cognitive mapping and spatial behavior. Aldine, Chicago, Ill.
- Dudnik, E.E. 1972. SECTION: Pages 928-931 in Tomlinson, R.F., ed., Geographical data handling. IGU Commission on Geographical Data Sensing and Processing, Ottawa, Ont.
- Dueker, K.J. 1966. Spatial data systems: Organization of spatial data, systems considerations, special topics. Technical Reports 4, 5, 6 of Office of Naval Research Task 389-143, Contract NONR 1228(37). Department of Geography, Northwestern University, Evanston, Ill.
- 1972. A framework for encoding spatial data. Geographical Analysis 4:98-105.

- Duncan, O. D., Cuzzort, R. P. and Duncan, B. 1961. Statistical geography: Problems in analyzing area data. The Free Press, Glencoe, Ill.
- Edson, D. T. 1973. United States Geological Survey, Department of the Interior, Reston, Va. Personal communication.
- Edwards, E. 1964. Information transmission. Chapman & Hall, New York, N. Y.
- Ekman, G., Lindman, R. and William-Olson, W. 1963. A psychophysical study of cartographic symbols. *Geografiska Annaler* 45:262-271.
- Electromagnetic Compatibility Analysis Center. 1969. SAM (Sight Analysis Model). Electromagnetic Compatibility Analysis Center, Annapolis, Md.
- Emerson, W. 1969. A review of geographic and psychologic research into the structure of environmental images. University of Minnesota, Minneapolis, Minn. Unpublished paper.
- Farnsworth, G. L. 1970. DIME applications and the computerized resource allocation model. American Statistical Association, Joint Statistical Meetings, Detroit, Mich.
- \_\_\_\_\_. 1973. Census Use Study, United States Bureau of the Census, Washington, D. C. Personal communication.
- Finkelstein, C. B. 1970. Data entry techniques. *Australian Computer Journal* 2:146-155.
- Fisher, H. T. 1970. The use of computer graphics in planning. Paper presented at the National Planning Conference of the American Society of Planning Officials.
- Flowerdew, R. 1973a. A logic of locational descriptions: Some illustrations from the context of geographical mobility. *Occasional Papers* 17. Department of Geography, University College, London, England.
- \_\_\_\_\_. 1973b. The use of formal spatial languages in geography. Paper presented to the Meeting of the Quantitative Methods Study Group, Institute of British Geographers, Newcastle. Mimeograph.
- \_\_\_\_\_. 1974. Geography Department, University College, London, England. Personal communication.
- Forrester, J. 1961. Industrial dynamics. M. I. T. Press, Cambridge, Mass.
- Fox, I. K. 1973. Planning and management of environment research programs. *Optimum* 4(4):38-49. (Publication of the Department of Supply and Services, Government of Canada, Ottawa, Ont.)
- Freeman, H. 1961. On the encoding of arbitrary geometric configurations. *Institute of Radio Engineers, Transactions on Electronic Computers* EC-10:260-268.
- \_\_\_\_\_. and Gardner, L. 1964. Apictorial jigsaw puzzles, the computer solution of a problem in pattern recognition. *Institute of Radio Engineers, Transactions on Electronic Computers* EC-13:118-127.

- Gaits, G. M. 1968. Mapping by computer: LINMAP. *The Cartographic Journal* 5(2):87.
- Garin, R. A. 1966. A matrix formulation of the Lowry model for intra-metropolitan activity location. *Journal of the American Institute of Planners* 32.
- Gould, P. 1965. On mental maps. Michigan Inter-University Community of Mathematical Geographers, Department of Geography, University of Michigan, Ann Arbor, Mich.
- \_\_\_\_\_. and White, R. 1974. Mental maps. Penguin Books, Baltimore, Md.
- Grigg, D. B. 1965. The logic of regional systems. *Annals of the Association of American Geographers* 55(3):465-491.
- \_\_\_\_\_. 1967. Regions, models and classes. Pages 461-509 in Chorley, R. J. and Haggett, P., eds., *Models in geography*. Methuen & Co., London, England.
- Hagerstrand, T. 1953. Innovationsförloppet ur korologisk synpunkt. Department of Geography, University of Lund, Lund, Sweden.
- \_\_\_\_\_. 1963. Geographical measurements in migration. In Sutter, J., ed., *Human displacements. Entretiens de Monaco en sciences humaines*, Monaco.
- \_\_\_\_\_. 1965. A Monte Carlo approach to diffusion. *European Journal of Sociology* 6:43-67.
- \_\_\_\_\_. 1967a. The computer and the geographer. *Transactions of the Institute of British Geographers* 42:1-19.
- \_\_\_\_\_. 1967b. Innovation diffusion as a spatial process. University of Chicago Press, Chicago, Ill.
- \_\_\_\_\_. 1970. Presidential address: What about people in regional science? *Papers of the Regional Science Association* 24:1-7.
- \_\_\_\_\_. Hannerberg, T. and Odeving, B., eds. 1957. Migration in Sweden: A symposium. Lund Studies in Geography, Series B 13. University of Lund, Lund, Sweden.
- Haggett, P. 1965. Locational analysis in human geography. Edward Arnold (Publishers), London, England.
- \_\_\_\_\_. 1969. On geographical research in a computer environment. *The Geographical Journal* 135(4).
- \_\_\_\_\_. and Chorley, R. F., eds. 1969. Network analysis in geography. Edward Arnold (Publishers), London, England.
- Haig, R. M. 1926. Towards an understanding of the metropolis. *Quarterly Journal of Economics* 40.
- Hall, P., ed. 1966. Von Thunen's isolated state. Pergamon Press, Oxford, England.
- Harbaugh, J. W. and Merriam, D. F. 1967. Computer applications in stratigraphic analysis. John Wiley & Sons, New York, N. Y.

- Hardy, E. E. and Shelton, R. L. 1970. Inventorying New York's land use and natural resources. New York's Food and Life Sciences 3(4).
- \_\_\_\_\_, Shelton, R. L. and Belcher, D. J. 1971a. New York State land use and natural resources inventory final report, volume I: Summary report. Center for Aerial Photographic Studies, Cornell University, Ithaca, N. Y.
- \_\_\_\_\_, Shelton, R. L., Belcher, D. J. and Roach, J. T. 1971b. New York State land use and natural resources inventory final report, volume II: Classification and inventory methods. Center for Aerial Photographic Studies, Cornell University, Ithaca, N. Y.
- Harvard University. 1968. Users reference manual for synagraphic computer mapping: SYMAP version V. Laboratory for Computer Graphics and Spatial Analysis, Harvard University, Cambridge, Mass.
- Harvey, D. 1969. Explanation in geography. St. Martin's Press, New York, N. Y.
- Healy, J. 1972. ABLES: Automated branch location and evaluation system. Papers from the 10th Annual Conference of the Urban and Regional Information Systems Association, San Francisco, Calif.
- Hebin, O. 1969. Computer drawn isarithmic maps. Geografisk Tidsskrift 68(2):50-63.
- Herbert, J. D. and Stevens, B. 1960. A model for the distribution of residential activity in urban areas. Journal of Regional Science 2:21-36.
- Hoelzel, J. 1973. Los Angeles County Department of Hospitals, Los Angeles, Calif. Personal communication.
- Holleb, Doris E. 1969. Social and economic information for urban planning. Center for Urban Studies, University of Chicago, Chicago, Ill.
- Holroyd, M. T. and Bhattachayya, B. K. 1970. Automatic contouring of geophysical data using bicubic spline interpolation. Paper 70-55. Geological Survey of Canada, Ottawa, Ont.
- Hoyt, H. 1933. One hundred years of land values in Chicago. University of Chicago Press, Chicago, Ill.
- \_\_\_\_\_. 1939. The structure and growth of residential neighborhoods in American cities. Federal Housing Administration, Washington, D. C.
- Huff, D. L. 1964. Defining and estimating a trading area. Journal of Marketing 28.
- Hurd, R. M. 1903. Principles of city land values. The Record and Guide, New York, N. Y.
- Ion, R. J. 1969. The geographic basis of the DBS geo-coding system for urban areas: An overview. Statistics Canada (Dominion Bureau of Statistics), Ottawa, Ont.
- Isard, W. 1956. Location and space-economy. John Wiley & Sons, New York, N. Y.

- Isard, W. 1970. On notions of models and time. *Papers of the Regional Science Association* 25:7-31.
- et al. 1960. *Methods of regional analysis: An introduction to regional science*. M.I.T. Press, Cambridge, Mass.
- and Reiner. 1966. *Regional science: Retrospect and prospect*. *Papers of the Regional Science Association* 16:1-16.
- Jaro, M. A. 1972a. OS ADMATCH: An address matching system. *Census Use Study Computer Program Package II*. U.S. Department of Commerce, Bureau of the Census, Washington, D.C.
- . 1972b. UNIMATCH: A computer system for generalized record linkage under conditions of uncertainty. *Proceedings of the American Federation of Information Processing Societies* 40. AFIPS Press, Montvale, N.J.
- Jarvis, C. L. 1971. A method for fitting polygons to figure boundary area. *Australian Computer Journal* 3:50-54.
- Johnson, T. E. 1963. SKETCHPAD III: A computer program for drawing in three dimensions. *Proceedings of the American Federation of Information Processing Societies* 23. AFIPS Press, Montvale, N.J.
- Joint Local Authority. 1972. *General information system for planning*. Report of the Joint Local Authority, Scottish Development Department and Department of Environment Study Team, London. Her Majesty's Stationery Office, London, Eng.
- Kain, J. F. 1962. The journey to work as a determinant of residential location. *Rand Paper P-2489*, Rand Corporation, Santa Monica, Calif.
- Kantorowich, R. H., et al. 1964. *Regional shopping centres: A planning report on north west England*. Department of Town and Country Planning, University of Manchester, Manchester, England.
- Kao, R. C. 1963. The use of computers in the processing and analysis of geographic information. *Geographical Review* 53:530-547.
- Kingston, P. P. C. H. 1968. A computer oriented geographic information processing system. Paper presented at the 10th Annual Conference of the Canadian Operational Research Society, Toronto, Ont.
- Klein, F. (English ed.) 1939. *Elementary mathematics from an advanced standpoint: Geometry*. New York, N.Y.
- Koeman, C. and van der Weiden, F. L. 1970. The application of computation and automatic drawing instruments to structural generalization. *The Cartographic Journal* 7:47-49.
- Kolars, J. F. and Nystuen, J. D. 1974. *Geography: The study of location, culture and environment*. McGraw-Hill, New York, N.Y.
- Koopmans, T. C. 1957. *Three essays on the state of economic science*. McGraw-Hill, New York, N.Y.
- Krasnakevitch, J. 1973. *Kindler Associates*, Harvard, Mass. Personal communication.

- Krumbein, W. C. 1959. Trend surface analysis of contour type maps with irregular control-point spacing. *Journal of Geophysical Research* 64(7):823-834.
- \_\_\_\_\_. 1966. Classification of map surfaces based on the structure of polynomial and Fournier coefficient matrices. *Kansas State Geological Survey Computer Contributions* 7.
- LaGarde, V. 1973. U.S. Army Engineers, Waterways Experiment Station, Vicksburg, Miss. Personal communication.
- Lang, T. 1971. Computer programs for mapping. *Architectural Press*, London, England.
- Leontieff, W. and Strout, A. 1963. Multiregional input-output analysis. In Barna, T., ed., *Structural interdependence and economic development*. Macmillan, London, England.
- Losch, A. 1954. *The economics of location*. Yale University Press, New Haven, Conn.
- Lowry, I. S. 1964. A model of a metropolis. RM-4035-RC. The Rand Corporation, Santa Monica, Calif.
- \_\_\_\_\_. 1965. A short course in model design. *Journal of the American Institute of Planners* 31:158-165.
- MacNeill, J. W. 1971. *Environmental management*. Information Canada, Ottawa, Ont.
- Matheron, P. 1970. The theory of regionalized variables. *Les cahiers du centre de morphologie mathématique de Fontainebleau* 5.
- McDonough, A. 1963. *Information, economics and management systems*. McGraw-Hill, New York, N. Y.
- McKay, G. A. 1973. Chief, Canadian Meteorological Service, Toronto, Ont. Personal communication.
- McKenzie, R. D. 1933. *The metropolitan community*. McGraw-Hill, New York, N. Y.
- McLeod, J. M. 1972. Personal communications and display of work at UNESCO/IGU Second Symposium on Geographical Information Systems, Ottawa, Ont.
- Meier, R. L. 1962. A communications theory of urban growth. Joint Center for Urban Studies. M.I.T. Press, Cambridge, Mass.
- Melander, C., Salomonsson, O. and Selander, K. 1971. FRIS A:2. A spatial information system - A pilot study. Data registration - Theoretical analysis of basic concepts. Central Board for Real Estate Data, Sundbyberg, Sweden.
- Merrill, R. D. 1973. Representation of contours and regions for efficient computer search. *Communications of the Association of Computing Machinery* 16(2):69-82.

- Meyer, M. A. 1972. The Census Bureau's approach to geographic base (DIME) file management. Papers from the 10th Annual Conference of the Urban and Regional Information Systems Association, San Francisco, Calif.
- Milgram, S. 1972. A psychological map of New York City. *American Scientist* 60(2):194-200.
- Ministry of Housing and Local Government, U.K. 1970. Information needs of the new planning system. Report of the Joint Local Authority, Scottish Development Department and Ministry of Housing and Local Government Study Team, London. Her Majesty's Stationery Office, London, England.
- Monkhouse, F. J. and Wilkinson, H. R. 1970. Maps and diagrams. (3rd ed.) Methuen & Co., London, England.
- Morrill, R. L. 1965. Migration and the spread and growth of urban settlement. Lund Studies in Geography Series B 26. Gleerup, Lund, Sweden.
- Murphy, E. M. 1972. The random rounding technique for guarding against illegal disclosure in published census tables. Statistics Canada, Ottawa, Ont.
- Murray, F. W. 1968. A method of objective contour construction. Report AD666830. United States Department of Commerce, Clearinghouse, Springfield, Va.
- Muth, R. T. 1961. The spatial structure of the housing market. Papers of the Regional Science Association 7:207-220.
- New York State. 1967. MAP01 manual. Department of Public Works, New York State, Albany, N. Y.
- Nourse, H. O. 1968. Regional economics. McGraw-Hill, New York, N. Y.
- Observatoire Économique Méditerranéen. 1971. Geocodification. No 203/1. Institut National de la Statistique et des Études Économiques, Marseille, France.
- Olsson, A. 1971. FRIS C:3. Programs for coordinate processing. KOSP - retrieval; PYTOM - calculation of areas and circumferences; RITA1 - plotting. Central Board for Real Estate Data, Sundbyberg, Sweden.
- and Selander, K. 1971. FRIS C:2. Dot maps by computer. Central Board for Real Estate Data, Sundbyberg, Sweden.
- Owen, E. H. 1972. An automated refuse collection system. Papers from the 10th Annual Conference of the Urban and Regional Information Systems Association, San Francisco, Calif.
- Pearson, A. K. 1972. LINMAP and COLMAP. Pages 955-958 in Tomlinson, R. F., ed., Geographical data handling. IGU Commission on Geographical Data Sensing and Processing, Ottawa, Ont.
- Petrie, G. 1966. Numerically controlled methods of automatic plotting and draughting. *The Cartographic Journal* 3(2):60-73.

- Peucker, T. C. 1972. Computer cartography: A working bibliography. Discussion Paper 12. Department of Geography, University of Toronto, Toronto, Ont.
- Pfaltz, J. L. 1967a. Manipulation and display of geographically distributed data. *Social Science Information* 11(4):217-222.
- . 1967b. MANS: A map analysis system. Technical report TR-67-42-NSG-398. Computer Science Center, University of Maryland, College Park, Md.
- . 1967c. Pattern recognition VI: MANS - A map analysis system. Technical report TR-67-42-NSG-398. Computer Science Center, University of Maryland, College Park, Md.
- and Rosenfeld, A. 1967. Computer representation of planar regions by their skeletons. *Communications of the Association of Computing Machinery* 10(2):119-123.
- Philbrick, A. K. 1967. The Erie project: Spatial terms of reference. Paper presented at the University of Western Ontario, Department of Geography Conference on the Erie Project. University of Western Ontario, London, Ont.
- Phillips, J. 1972a. TARELA II: A tabulation request language: User's manual. General Survey Systems, Statistics Canada, Ottawa, Ont.
- . 1972b. Safeguarding against disclosure in STATPAK version 3. General Survey Systems, Statistics Canada, Ottawa, Ont.
- Phillips, R. L. 1972. Interactive mapping for environmental systems. Aerospace Engineering Department, University of Michigan, Ann Arbor, Mich. Mimeograph.
- Pierce, J. R. 1961. Symbols, signals and noise: The nature and process of communication. Harper & Row, New York, N. Y.
- Pitts, F. 1963. Problems in computer simulation of diffusion. *Papers of the Regional Science Association* 11:111-119.
- Podehl, W. M. 1972a. STATPAK and its data base: A guide for creating a STATPAK data base. General Survey Systems, Statistics Canada, Ottawa, Ont.
- . 1972b. STATPAK (version 3) general information manual. General Survey Systems, Statistics Canada, Ottawa, Ont.
- . 1973. General Survey Systems, Statistics Canada, Ottawa, Ont. Personal communication.
- Porter, J. 1965. The vertical mosaic: An analysis of social class and power in Canada. University of Toronto Press, Toronto, Ont.
- P. R. C. Information Sciences Company. 1973. Lineal input system users manual. Published for Air Force Systems Command, Griffiss Air Force Base, Rome, N. Y.
- Quastler, H. 1956. Studies of human channel capacity. In Cherry, E. C., ed., *Information theory. Proceedings of the Third London Symposium*, 1955. Butterworth, London, England.



- Rayner, J. N. 1971. An introduction to spectral analysis. Pion, London, England.
- \_\_\_\_\_. 1974. The time element in spatial process. Paper presented at the Symposium on Geographic Analysis and Understanding Spatial Process. Department of Geography, University of Michigan, Ann Arbor, Mich.
- Reilly, W. J. 1929. Methods for the study of retail relationships. Bulletin 2994, Research Monograph 4. University of Texas, Austin, Texas.
- Rhind, D. W. 1971. Towards instant and efficient maps: The work of the Experimental Cartography Unit. Review of Geographical Monitoring 25(4):391-398.
- Richardson, H. W. 1969. Elements of regional economics. Penguin Books, Harmondsworth, Middlesex.
- Robinson, A. H. 1956. The necessity of weighting values in correlation analysis of area data. Annals of the Association of American Geographers 46:233-236.
- \_\_\_\_\_. 1969. Elements of cartography (3rd ed.). John Wiley & Sons, New York, N. Y.
- Rogers, A. 1968. Matrix analysis of interregional population growth and distribution. University of California Press, Berkeley, Calif.
- Rushton, G., Horton, F. E., Lindberg, J. B., McNulty, M. L., Reynolds, D. R., Salisbury, N. E., and Taylor, P. 1971. Quantitative geography: Achievements and prospects. Final Report, ETL-CR-71-12. U. S. Army Engineer Topographic Laboratories, Research Institute, Department of the Army, Fort Belvoir, Va.
- Salmona, J. 1974. Opération Pilote Interministérielle sur les Données Administratives, Marseille, France. Personal communication.
- Salomonsson, O. 1969. Data banking systems for urban planning. Proceedings of the Conference on Information and Planning 2. CES-IP-8. Centre for Environmental Studies, London, England.
- \_\_\_\_\_. 1973. NORDPLAN, Skeppsholmen, S111 49 Stockholm, Sweden. Personal communication.
- Schmidt, A. 1966. SYMAP, a users manual. Tri-County Regional Planning Commission, Lansing, Mich.
- \_\_\_\_\_. 1970. Selected projects 1966-1970. Laboratory for Computer Graphics and Spatial Analysis, Harvard University, Cambridge, Mass.
- \_\_\_\_\_. 1972a. SYMAP. Pages 934-938 in Tomlinson, R. F., ed., Geographical data handling. IGU Commission on Geographical Data Sensing and Processing, Ottawa, Ont.
- \_\_\_\_\_. 1972b. SYMVU. Pages 1041-1044 in Tomlinson, R. F., ed., Geographical data handling. IGU Commission on Geographical Data Sensing and Processing, Ottawa, Ont.

- Schmidt, W. E. 1969. The AUTOMAP system. *Surveying and Mapping* 29(1):101-106.
- \_\_\_\_\_. 1970. Cartographic automatic mapping program (CAM). Documentation accession No. PB191779. U.S. Department of Commerce, Clearinghouse, Springfield, Va.
- \_\_\_\_\_. 1973. Central Intelligence Agency, Washington, D. C. Personal communication.
- Science Council of Canada. 1969. Science and technical information in Canada: Part II - Government departments and agencies. Special Study 8. Information Canada, Ottawa, Ont.
- Seeman, I. 1973. Personal communication.
- Selander, K. 1970. FRIS C:1. Registration and storing of coordinates. Central Board for Real Estate Data, Sundbyberg, Sweden.
- Shelton, R. L. and Dimock, R. A. 1971a. New York State land use and natural resources inventory final report, volume IV: PLANMAP II and DATALIST I user manuals. Center for Aerial Photographic Studies, Cornell University, Ithaca, N. Y.
- \_\_\_\_\_. 1971b. New York State land use and natural resources inventory final report, volume V: Computer programs. Center for Aerial Photographic studies, Cornell University, Ithaca, N. Y.
- Shelton, R. L., Dimock, R. A. and Hope, J. M. 1971. New York State land use and natural resources inventory final report, volume III: Data system and data processing. Center for Aerial Photographic Studies, Cornell University, Ithaca, N. Y.
- Singer, H. W. 1936. The 'Courbe des populations': A parallel to Pareto's law. *Economic Journal* 46.
- Sinton, D. and Steinitz, C. 1969. GRID: A users reference manual. Laboratory for Computer Graphics and Spatial Analysis, Harvard University, Cambridge, Mass.
- Solosko, R. B. 1973. A semi-automatic system for digitizing bathymetric charts. Calspan report VF-5206-XI. Final report. Contract N00014-72-C0466. Prepared for the Long Range Acoustic Propagation Project, Office of Naval Research. Department of the Navy, Washington, D. C.
- Statistics Canada. 1972a. GRDSR: Facts by small areas - An introduction. Methodology and Systems Branch, Statistics Canada, Ottawa, Ont.
- \_\_\_\_\_. 1972b. GRDSR: Geographically referenced data storage and retrieval system - A technical description. Methodology and Systems Branch, Statistics Canada, Ottawa, Ont.
- \_\_\_\_\_. 1972c. GRDSR: Guidelines for requesting geographical areas. Census User Inquiry Service, Census Division, Statistics Canada, Ottawa, Ont.

- Stea, D. 1969. The measurement of mental maps: An experimental model for studying conceptual spaces. *Studies in Geography 17: Behavioral problems in geography: A symposium*. Northwestern University, Evanston, Ill.
- Steiner, D. 1970. GEOMAP. Personnel communication, University of Waterloo, Waterloo, Ont.
- Sutherland, I.E. 1963. SKETCHPAD: A man-machine graphical communication system. *Proceedings of the American Federation of Information Processing Societies 23:324-346*. AFIPS Press, Montvale, N.J.
- Switzer, P. 1973. Estimation of the accuracy of qualitative maps. Paper presented at the NATO Advanced Study Institute Conference on Display and Analysis of Spatial Data, Nottingham, England.
- Switzer, W. 1973. Land Management Information Systems, Lands Directorate, Environment Canada, Ottawa, Ont.
- Symons, D.C. 1972a. Automatic mapping and its relation to urban information systems. *The Canadian Surveyor 26(5):545-557*.
- \_\_\_\_\_. 1972b. Parcel geocoding for urban and rural information. In *Geocoding '72. Papers of the Geographic Base File Special Interest Group, 10th Annual Conference, Urban and Regional Information Systems Association, San Francisco, Calif.*
- Tarrant, J.R. 1968. Computers in geography. *Geo Abstracts, School of Environmental Studies, University of East Anglia, Norwich, England*.
- Taylor, L.R. 1973. Comprehensive Planning Organization, San Diego County, San Diego, Calif. Personal communication.
- Thomas, A. 1972. OBLIX. Department of Geography, University of Edinburgh, Edinburgh, Scotland. Personal communication.
- Thomas, E.N. and Anderson, D.L. 1965. Additional comments on weighting values in correlation analysis of areal data. *Annals of the Association of American Geographers 55:492-505*.
- Thompson, D.R. 1967. A cartographic scanner. *Proceedings of the 27th Annual Meeting of the American Society of Photogrammetry - American Congress on Surveying and Mapping, Washington, D.C.*
- Tobin, J. 1953. Research on full employment and inflation. Unpublished memorandum prepared for the Social Science Research Council, U.S.A.
- Tobler, W.R. 1962. A classification of map projections. *Annals of the Association of American Geographers 52:196-275*.
- \_\_\_\_\_. 1963. Geographic area and map projections. *Geographical Review 53:59-78*.
- \_\_\_\_\_. 1964a. Automation in the preparation of thematic maps. *Journal of the British Cartographic Society, June 1964:1-7*.
- \_\_\_\_\_. 1964b. Computation of the correspondence of geographical patterns. *Papers of the Regional Science Association 15:131-139*.

- Tobler, W.R. 1965. Automation in the preparation of thematic maps. *Cartographic Journal* 2(1):32-38.
- \_\_\_\_\_. 1966. Numerical map generalization and notes on the analysis of geographical distributions. Discussion Paper 8. Michigan Inter-University Community of Mathematical Geographers. Department of Geography, University of Michigan, Ann Arbor, Mich.
- \_\_\_\_\_, ed. 1970. Selected computer programs. Department of Geography, University of Michigan, Ann Arbor, Mich.
- \_\_\_\_\_, et al. 1968. The philosophy of maps. Discussion Paper 12. Michigan Inter-University Community of Mathematical Geographers, Department of Geography, University of Michigan, Ann Arbor, Mich.
- Tomlinson, K.R. and Huff, A. 1972. The creation of geo-coded files, aerial retrieval and mapping. Proceedings of the Planning and Transport Research and Computation Co. Ltd. Symposium on Data Management, London, England.
- Tomlinson, R.F. 1962. Computer mapping: An introduction to the use of electronic computers in the storage, compilation and assessment of natural and economic data for the evaluation of marginal lands. Report presented to the National Land Capability Inventory Seminar held under the direction of the Agricultural Rehabilitation and Development Administration, Canada Department of Agriculture, Ottawa, Ont.
- \_\_\_\_\_. 1963. Feasibility report of computer mapping system. Spartan Air Services Ltd., Ottawa, Ont.
- \_\_\_\_\_. 1967. An introduction to the geographic information system of the Canada Land Inventory. Department of Forestry and Rural Development, Ottawa, Ont.
- \_\_\_\_\_. 1968. A geographic information system for regional planning. Pages 200-210 in Stewart, G.A., ed., Land evaluation. Papers of a CSIRO Symposium organized in cooperation with UNESCO. Macmillan of Australia, South Melbourne, Victoria.
- \_\_\_\_\_. 1971. The economic, sociological and political nature of data banks for development. Pages 20-31 in Data banks for development. Proceedings of the International Expert Meeting on Data Banks for Development, Saint-Maximin, France. I.N.S.E.E., Observatoire Economique Méditerranéen, Marseille, France.
- \_\_\_\_\_, ed. 1972. Geographical data handling (Symposium ed.). UNESCO/IGU Second Symposium on Geographical Information Systems, Ottawa, Ont. IGU Commission on Geographical Data Sensing and Processing, Ottawa, Ont.
- Treichel, R. 1973. National Civil Defense Computer Facility, Olney, Md. Personal communication.
- Trowbridge, C.C. 1913. On fundamental methods of orientation and imaginary maps. *Science* 38.

- Unidata, Inc. 1973. INSIGHT users guide. Unidata, Inc., Ann Arbor, Mich.
- University of Illinois. 1970. NARIS: The natural resources information system. Illiac IV project, in cooperation with the Northeast Illinois Natural Resource Service Center, Urbana, Ill.
- U.S. Bureau of Census. 1970a. Census use study: Data uses in health planning. Report 8. U.S. Department of Commerce, Bureau of the Census, Washington, D.C.
- \_\_\_\_\_. 1970b. Census use study report series, volumes 1 to 13. U.S. Department of Commerce, Bureau of the Census, Washington, D.C.
- \_\_\_\_\_. 1971a. Use of address coding guides in geographic coding: Case studies. Report GE60 No. 1. U.S. Government Printing Office, Washington, D.C.
- \_\_\_\_\_. 1971b. Census use study: geocoding with ADMATCH - A Los Angeles experience. Report 14. U.S. Department of Commerce, Bureau of the Census, Washington, D.C.
- \_\_\_\_\_. 1971c. Census use study: Health information system II. Report 12. U.S. Department of Commerce, Bureau of the Census, Washington, D.C.
- \_\_\_\_\_. 1971d. Census use study: Computer resource allocation model (CRAM). Report 13. U.S. Department of Commerce, Bureau of the Census, Washington, D.C.
- \_\_\_\_\_. 1972. DIME: A geographic base file package. U.S. Department of Commerce, Bureau of the Census, Washington, D.C.
- \_\_\_\_\_. 1973. Census use study: DIME workshops - An interim report. U.S. Department of Commerce, Bureau of the Census, Washington, D.C.
- U.S. Dept. of Commerce. 1967. A program of composite economic and geographic analysis using a computer mapping system. U.S. Department of Commerce, Office of Regional Development Planning, Washington, D.C.
- U.S. Geological Survey (U.S.G.S.) 1969. Proceedings of the symposium on map and chart digitizing, October 1969. Geological Survey Computer Contribution 5. Topographic Division, Geological Survey, U.S. Department of the Interior, Washington, D.C.
- Vogt, Ivers and Associates. 1966. Comprehensive land use inventory procedures manual for data collection and analysis. Commonwealth of Massachusetts, Boston Regional Planning Project, Boston, Mass.
- Von Thunen, J.H. 1826. *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*. Hamburg. (Translation in Hall, P., ed., 1966. Von Thunen's isolated state. Pergamon Press, Oxford, England.
- Ward, N. 1962. *The public purse: A study in Canadian democracy*. University of Toronto Press, Toronto, Ont.
- Wärneryd, O. 1968. *Interdependence in urban systems*. Regionkonsult Aktiebolag, Göteborg, Sweden.

- Warntz, W. 1971. Computer graphics and theoretical geography. In Research Institute lectures on geography, special report ETL SR 71 1. U.S. Army Engineer Topographic Laboratories, Fort Belvoir, Va.
- Waugh, T. C. 1972. The choropleth mapping system. Inter-University Research Councils Series Report 7. Edinburgh Catalogue No. 18,900,103. Scientific and Social Sciences Program Library, University of Edinburgh, Edinburgh, Scotland.
- Webber, M. M. 1963. Order in diversity: Community without propinquity. In Wingo, L., ed., Cities and space: The future use of urban land. Johns Hopkins Press, Baltimore, Md.
- . 1964. The urban place and nonplace urban realm. In Webber, M. M. et al. Explorations into urban structure. University of Pennsylvania Press, Philadelphia, Pa.
- Weiner, N. 1930. Generalised harmonic analysis. *Acta Mathematica* 55:117-258. (Reprinted M.I.T. Press 51, 1964.)
- . 1958. Non-linear problems in random theory. M.I.T. Press, Cambridge, Mass.
- Westerfield, F. 1973. Regional Planning Council, Baltimore, Md. Personal communication.
- White, M. S. 1973. Census Use Study, U.S. Department of Commerce, Bureau of the Census, Washington, D.C. Personal communication.
- White, R. R. 1970. Geographic information and the interpretation of choropleth maps. Ph.D. Dissertation, Department of Geography, University of Bristol, Bristol, England.
- Williams, C. M. 1973. Computer Science Department, Pennsylvania State University, Philadelphia, Pa.
- Williams, R. 1968. Data structures: An annotated bibliography. Department of Electrical Engineering, New York University, New York, N.Y.
- . 1971. A survey of data structures for computer graphics systems. *Computing Surveys* 3(1):1-22.
- Willis, J. 1971. Design issues for urban and regional information systems. Working paper CES-WP-71. Centre for Environmental Studies, London, England.
- Wilson, A. G. 1967. A statistical theory of spatial distribution models. *Transportation Research* 1:253-269.
- . 1968. Interregional commodity flows: Entropy maximising approaches. Working paper CES-WP-19. Centre for Environmental Studies, London, England.
- . 1969a. Development of some elementary residential location models. *Journal of Regional Science* 9(3):377-385.
- . 1969b. Entropy in urban and regional modelling. Working paper CES-WP-26. Centre for Environmental Studies, London, England.
- . 1969c. Disaggregating elementary residential location models. Working paper CES-WP-37, Centre for Environmental Studies, London, England.

- Wilson, N.L. 1955. Space, time and individuals. *Journal of Philosophy* 52(22):589-598.
- Wingo, L. Jr. 1961a. An economic model of the utilization of urban land for residential purposes. *Papers of the Regional Science Association* 7:191-205.
- \_\_\_\_\_. 1961b. *Transportation and urban land*. Johns Hopkins Press, Baltimore, Md.
- Zipf, G.K. 1949. *Human behavior and the principle of least effort*. Addison-Wesley Publishing Co., Reading, Mass.

APPENDIX     TECHNICAL DESCRIPTION OF THE CANADA  
GEOGRAPHIC INFORMATION SYSTEM



## APPENDIX - TABLE OF CONTENTS

	Page
Table of Contents .....	307
List of Figures .....	311
List of Tables .....	314
<u>Data Input Subsystem</u> .....	315
<u>Data Reduction Subsystem and Manual Error Correction Subsystem</u> ..	317
Phase 0 .....	322
Phase 1 .....	322
Initialization and control .....	325
Compute Section size and get first section .....	325
Compute V-values .....	325
Vertex table initialization .....	326
Cloud elimination .....	326
Boundary tracing and colouring .....	328
Write segment record and get next segment .....	329
Write equivalent colours .....	329
Corner point location .....	329
Error routine .....	330
Update of map control block .....	330
Phase 2 .....	331
Initialization and control .....	334
Location of map corners .....	335
Compute transformation of coefficients .....	335
Transform descriptor-related reference coordinates to GCS .....	335
Write out revised descriptor data records and section records .....	335
Error analysis .....	336
Write map control block and end of phase .....	336
Phase 3 .....	337
Initialization and control .....	340
Framing .....	340
Colour equivalence resolution .....	344
Linkage .....	345
Transformations .....	346
Resolution reduction - scale change .....	346
Gap closing .....	347
Single point smoothing .....	347
Compacting .....	348
Geodetic calculations .....	350
Error tracing .....	352

	Page
Phase 4A.....	354
Coverage initialization .....	355
Initialize and process frame records.....	355
Read colour records .....	357
Processing the frame border.....	357
Bypass small noise faces.....	358
Trace boundary, concatenate segments.....	361
Error check.....	363
Completion of area centroid calculations .....	363
Build IDS records.....	364
Update MCB and write DDS .....	364
Elimination of small faces and lines .....	365
Manual Error Correction (MEC) subsystem.....	366
Error listings.....	366
Determine corrections, optional plot .....	368
Error correction program (MECMAIN).....	369
Phase 4B.....	371
Initialization .....	372
Read new IDS and IDS master and identify border records	372
Phase 5.....	373
Initialization .....	373
Read input and store border record - calculate	
geodetics .....	374
Determine equivalence .....	374
Write face and border data.....	375
Phase 6.....	375
Initialization .....	378
Read face records from phase 5. Store face numbers	
and equivalences in table. Accumulate geodetic data .....	378
Read map border file for face. Working file of	
frame borders.....	378
Read DDS master for face .....	378
Determine equivalence within face.....	378
Write delete instructions for DDS master.....	379
Compare adjacent descriptors. Assign face numbers	
if equivalent.....	379
Write out IDS and DDS face number change instructions..	379
Close DDS master file .....	379
Read work file of frame borders. Write out frame	
border delete and change instructions.....	379
Phase 7.....	380
Initialization .....	381
Read input files for each face.....	381
Edit files for validity on a face group basis.....	381
Recalculate geodetic information to be merged.....	381
Merge.....	382
Write revised DDS master.....	382
Update MCB .....	382

	Page
Phase 8 .....	382
Initialization, check sequence, and edit for validity.....	384
Read frame record. Read face and segment record and map border equivalence record.....	384
Compare segment length.....	384
Split segment.....	384
Join multiple exterior faces .....	384
Copy and remove deleted records. Write out revised IDS .....	385
End program.....	385
<u>Data Retrieval Subsystem</u> .....	387
<u>Command Language Interpreter</u> .....	391
Source program library (SYMLIB) .....	391
Routine program library (RTNLIB).....	391
Loadable library (LODLIB).....	391
Examples of requests.....	393
<u>Command Languages</u> .....	395
Manipulation language - #ASSESS command .....	395
Assessment language.....	396
Manipulation language.....	397
#OVERLAY .....	399
<u>Plot</u> .....	407
<u>Examples of Output from the Canada Geographic Information System</u> .....	407
Example 1 - Area totals from maps of single data type.....	408
Example 2 - Overlay and comparison of maps of two different data types.....	409
Example 3 - Identification of land characteristics adjacent to specific feature. Production of graphic output .....	414
Example 4 - Overlay and comparison of maps of two different data types.....	419
Example 5 - Multiple overlays. Weighted comparisons. List and plotted output of selected areas. ....	421
<u>Examples of V-value Allocation, Cloud Elimination, and Boundary Tracing</u> .....	438

## Page

<u>Partial List of Persons Instrumental in the Development of the Canada Geographic Information System in Approximate Order of Appearance</u> .....	443
<u>Canada Geographic Information System Bibliography</u> .....	444

## APPENDIX - LIST OF FIGURES

Fig.	Title	Page
A. 1a	Data reduction subsystem flow chart	319
A. 1b	Data reduction subsystem flow chart (continued)	320
A. 2	Phase 1 inputs and outputs	323
A. 3	Phase 1 flow chart	324
A. 4	Computation of V-values	326
A. 5	Direction of line-following scan ("d" code)	327
A. 6	Direction of second half of line-following scan	328
A. 7	Creation of gap in output image	331
A. 8	Relationship of areas in two coordinate systems	332
A. 9	Conceptual functions of phase 2	333
A. 10	Phase 2 flow chart	334
A. 11	Inputs and outputs of phase 3	337
A. 12	Phase 3 flow chart	338
A. 13	Calling sequence of routines in phase 3	339
A. 14	Relationship of unit grids to unit frames	341
A. 15	Sequence of frame numbers	342
A. 16	Morton matrix	343
A. 17	Phase 3 operations	344
A. 18	Colour equivalence resolution	344
A. 19	Linkage process	345
A. 20	Transformations	346
A. 21	Gap closing	347
A. 22	Single point smoothing	347
A. 23	Two-point smoothing or generalization	348
A. 24	Compacting	350
A. 25	Level 3 measurement area	351
A. 26	Structure of IDS and DDS	354
A. 27	Phase 4A flow chart	356
A. 28	Small noise face (interior)	359
A. 29	Small noise face (exterior)	360

Fig.	Title	Page
A. 30	Elimination of vertices at section borders .....	362
A. 31	Exterior and interior face boundaries of colour 8 .....	362
A. 32	A compound error .....	365
A. 33	Distortion error in phase 4 .....	365
A. 34	MEC subsystem flow chart .....	367
A. 35	Phase 4B flow chart .....	372
A. 36	Phase 5 flow chart .....	373
A. 37	Concept of matching .....	374
A. 38	Non-matching vertices and descriptors .....	374
A. 39	Frame-aligned segment .....	375
A. 40	Phase 6 flow chart .....	377
A. 41	Equivalences within face .....	378
A. 42	Deletions in DDS master file .....	379
A. 43	Matching over frame and map boundaries .....	380
A. 44	Phase 7 flow chart .....	381
A. 45	Phase 8 flow chart .....	383
A. 46	Joining multiple exterior faces .....	385
A. 47	Sample of IDS format .....	386
A. 48	CGIS retrieval monitor system .....	389
A. 49	Conceptual diagram of #OVERLAY command functions .....	400
A. 50	Flow chart showing sequence of logical operations in IDS overlay .....	401
A. 51	Flow chart showing sequence of logical operations in DDS overlay .....	404
A. 52	OVQ process .....	405
A. 53	Sorting of face equivalence record .....	406
A. 54	Example 3 - CGIS graphic output .....	418
A. 55	Recreation 605 - Original manuscript copy .....	424
A. 56	Agriculture 205 - Original manuscript copy .....	425
A. 57	Forestry 305 - Original manuscript copy .....	426
A. 58	Wildlife Ungulates 405 - Original manuscript copy .....	427
A. 59	Wildlife Waterfowl 505 - Original manuscript copy .....	428
A. 60	Example 5 - CGIS graphic output .....	429

Fig.	Title	Page
A. 61	Input map	438
A. 62	Scanner output	439
A. 63	Result of V-value calculations	440
A. 64	Result of cloud elimination - first line	441
A. 65	Result of cloud elimination - remaining lines	442

## APPENDIX - LIST OF TABLES

Table	Title	Page
A. 1	Canada Geographic Information System - example 1.....	408
A. 2a	Canada Geographic Information System - example 2.....	410
A. 2b	CGIS - example 2, continued .....	411
A. 2c	CGIS - example 2, continued .....	412
A. 2d	CGIS - example 2, continued .....	413
A. 3a	Canada Geographic Information System - example 3.....	415
A. 3b	CGIS - example 3, continued .....	416
A. 3c	CGIS - example 3, continued .....	417
A. 4	Canada Geographic Information System - example 4.....	420
A. 5	Weighting factors - examples.....	422
A. 6a	Canada Geographic Information System - example 5.....	430
A. 6b	CGIS - example 5, continued .....	431
A. 6c	CGIS - example 5, continued .....	432
A. 6d	CGIS - example 5, continued .....	433
A. 6e	CGIS - example 5, continued .....	434
A. 6f	CGIS - example 5, continued .....	435
A. 6g	CGIS - example 5, continued .....	436
A. 6h	CGIS - example 5, continued .....	437



APPENDIX - TECHNICAL DESCRIPTION OF THE CANADA  
GEOGRAPHIC INFORMATION SYSTEM (CGIS)

This technical description is intended to allow the reader fully to understand the functions of the Canada Geographic Information System, to present examples of its products, to describe the operations involved in achieving each function, and to give details of the algorithms underlying those operations. The system is currently in operation at the Department of the Environment, Government of Canada, Ottawa. Detailed system documentation and the computer programs reside with the Department of the Environment. Research leading to the development of the system started in 1961. Since that time more than 40 persons have been active in its development. Full acknowledgement of their efforts is given at the end of this technical description and in the text. The system is considered to be operational in terms of the functions described in this appendix in that it generates substantial products for administrative purposes, for various levels of government in Canada. The system facilities are, however, under continuing development.

DATA INPUT SUBSYSTEM

The data input subsystem converts the graphic source data to a simple non-graphic format. Three types of data are extracted from each source map: the image data, the descriptor data, and the linkage data.

The image data are the lines that form the boundaries on the outer edge of the map, plus all the line images contained in its interior. This web of graphic data is traced (scribed) on a clean sheet.

The descriptor data are written descriptions of the elements that comprise the map, such as descriptions of the present land use of each individual area on a present-land-use map. Each descriptor for each individual map element is sequentially numbered and listed on a form designed for the purpose.

The third type of data is the key to linking the image data to the descriptor data. A transparent sheet is overlaid on the source map and the descriptor list number is written at the approximate centre of the map element to which each descriptor relates.

These three source map derivatives are passed through a drum scanner, an X-Y digitizer, and an encoder, respectively, and are later put on magnetic tape to form the input to the data reduction subsystem. Diagram 5.11 in Chapter 5 above illustrates the data preparation process.

The traced sheet of image data is placed on the drum scanner and the scanning operation produces a digitized map of the image on magnetic tape. The drum scanner was developed to meet Canada Geographic Information System requirements by the International Business Machines Company (IBM). Use of the drum scanning approach was first considered in 1963, design criteria were established by the author in 1964, and development work was contracted to the following year. The scanner consists of a cylindrical drum on which a map or chart can be mounted, and a movable carriage that slowly moves the scanning head across the front of the revolving drum. The scanning system consists of the scanning head, its associated electronics, and controls leading to a standard IBM magnetic tape drive. The technique employed is to detect changes in intensity of light reflected from black or white areas on the map surface, and to record this information as a series of binary bits written on magnetic tape. The scan head uses fibre optics and can scan eight scan lines simultaneously. The drum scanner can accept a map of up to 48 in. x 48 in. A full-sized map takes approximately 15 minutes to scan, including the time to mount and dismount it, and smaller sheets take a correspondingly shorter time.

It is not within the scope of this appendix to give a detailed description of the drum scanner; the engineering aspects of the design have been described by Thompson.<sup>1</sup> However, the format of the map image data on tape is pertinent. One map-image record is produced for each 0.032 in. along the X axis of the map sheet, and the height of each record is 0.004 in. along the Y axis. The 0.032-in. record, comprising one byte of computer storage, is divided into eight bits. Each bit thus represents an area or spot 0.004 in. wide. Lines drawn on the map are usually 0.008 in. wide. The advantage of scribing as opposed to tracing by pencil or pen is that the minimum line width can be assured without excessive line thickness. If the scan heads identify 50% or more of a spot as part of a line, then the bit generated is 1; if not, the bit generated is 0. In this manner, a line is represented as

1. Thompson 1967.

a collection of 1-bits, which are usually 1, 2, or 3 spots in width.

The scribed image sheet, with the transparent numbered overlay superimposed, is placed on a D-mac cartographic X-Y digitizer with the four corner reference points and the coordinates of one reference point per map "face" coded in digits. A map face is any one of the distinct areas that together make up the surface of the map.

The magnetic tape from the X-Y digitizer is then placed on an NCR encoder and the sequentially numbered descriptor data are encoded beside the appropriate reference coordinates. These two processes are duplicated to allow for validity checks in the first part of the data reduction subsystem.

#### DATA REDUCTION SUBSYSTEM AND MANUAL ERROR CORRECTION SUBSYSTEM

The data reduction subsystem accepts the records produced by the scanner, digitizer, and encoder and, in concert with the manual error correction (MEC) subsystem, transforms them into a compact, error-free format that is convenient for storage and subsequent computer processing.

"Logical operations" are the steps taken to carry out the work; in the data reduction subsystem they are grouped into "phases". Each phase is a set of generally related logical operations, and has a set of inputs that it operates upon or refers to and a set of outputs that it produces for use by subsequent phases. These phases and logical operations are the building blocks with which it is possible to describe a system. The computer itself responds to a series of programs, made up of routines and subroutines, which work with its own resident operating system and take advantage of its specific computer architecture. Occasionally a logical operation is handled by a single program, but normally it requires the use of several programs, routines, or subroutines, some or all of which may also be used for other logical operations. In 1969, the actual program description for the Canada Geographic Information System required a substantial amount (19 volumes) of documentation. Since that time, programs have been modified and further programs have been written. The system continues to be under active development. Accordingly, this part of the appendix will concern itself with briefly describing the subsystem in terms of the logical operations carried

out by each phase. The names of programs and routines involved will be given where appropriate.

A flow chart of the data reduction subsystem is presented below as a guide to the flow of data through the phases. Although it provides only a brief indication of the operations within each phase, it gives a picture of the data that enter each phase, the data produced by each phase, and the linkages between phases; it may thus offer a reference point for the descriptions of individual phases.

In general, phases are linked by "sorts", by which the phase output is rearranged into a convenient format for the next phase. Each phase has some way of reporting errors. In phases 1 to 4A, the types of errors to be reported are specified on a "diagnostic card". The errors are traced and records of them are printed out at the end of each phase on a "system messages" list. Special procedures are built into the programs to report upon and resolve data errors.

As each map is processed through the system, vital statistics about its size, complexity, error state, and so forth are listed on a map control block (MCB). This is passed from phase to phase rather like a notebook. Each phase uses the data it contains to prepare the appropriate-sized files, tables, or switches for processing, or extracts map constants from it for use in calculating changes of map projection or scale, and so on. Toward the end of the work, in phase 6, the processing of individual maps gives way to the processing of all maps as one unit. At this point the MCB is joined by the master coverage control block (MCCB), which acts as the notebook for all the maps in the data bank of one data type (for example, land use, census, or recreation potential). The MCCB records are used in the same manner as the MCB records, but refer to the entire data bank in its final format.

It can be seen from the flow chart that in phase 0 the data are accepted from the digitizer and encoder and edited for validity and consistency. These data are the numbered records that describe the map elements, and the similarly numbered reference coordinates that allow each description to be related to a place on the map.



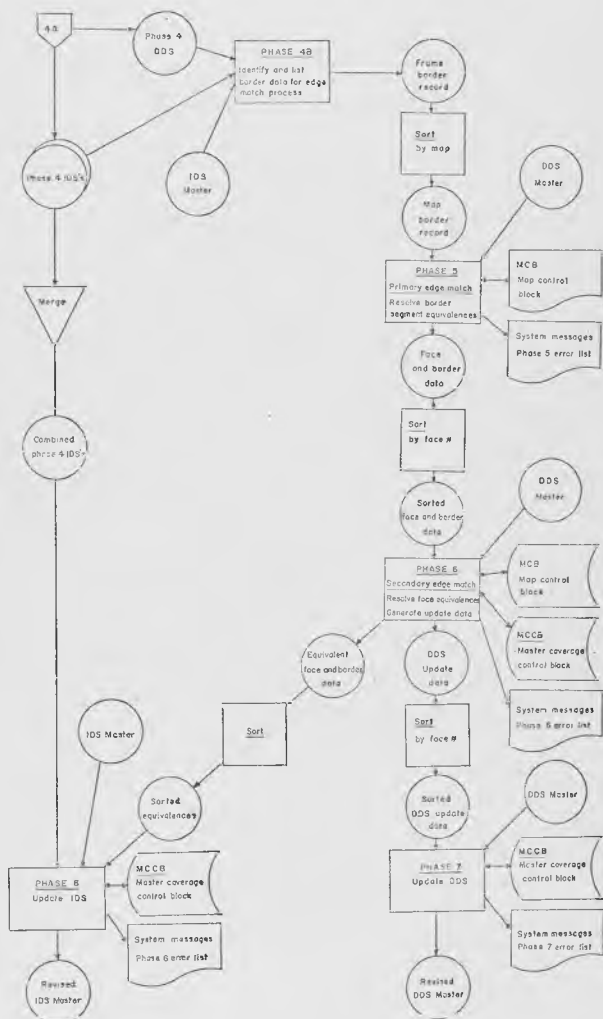


Fig. A. 1b. Data reduction subsystem flow chart (continued).

In phase 1, the raw image data are accepted directly from the scanner. This large amount of data is examined piece by piece, the graphic elements (lines, intersections, map borders, etc.) that make up the map are identified and labelled, and the rest is rejected.

Phase 2 prepares the way for bringing the descriptor data and the image data together. Because they were produced on different input devices, they arrive for processing with inherently different frames of reference. The coordinate system of the scanner data must thus be put into a form that can be related to the coordinate system of the data from the digitizer, specifically the reference coordinates and their related pieces of descriptor data. The transformations necessary to achieve this are calculated in phase 2.

Phase 3 brings the data together. In effect, this is the first time in the subsystem that the source map is recreated. Once the data are assembled in phase 3, distortions can be eliminated and the records compacted to an economical level for further handling. Phase 4A is another major phase. It takes the combined records from phase 3 and from them assembles the image file or "image data set" (IDS), and the descriptor file or "descriptor data set" (DDS), in the same format as that of the final data bank.

The different data records had to be brought together in phase 3 so that the image and descriptor could be matched properly. Now it is convenient to separate them, because the retrieval processes that eventually operate on the image file need not be burdened with the descriptor data, and similarly the retrieval processes that eventually operate on the descriptor data work immeasurably faster if they are not required to retain the image data. Pointers from one file to another allow joint processing as desired. Phase 4A therefore involves building linkages between the components of the map, and allows a detailed topological check to be made of the map components and a listing of errors to be put out for correction. These processes are handled by the manual error correction subsystem, which recycles the corrections through phase 4A until a clean set of new map data is achieved.

Phases 4B to 8 are involved with aligning the edges of the new maps being processed with those already in the data bank (the edge-match problem). Phase 4B is to identify and list the records that fall on the borders and must

be matched. Phases 5 and 6 are a two-stage edge matching process; phase 5 operations make sure that the descriptors on both sides of the interface are aligned, and the phase 6 process traces the entire boundaries of single areas that may fall across the interface, and makes sure that they have the same descriptor. A set of instructions is also prepared for changes that have to be made to both the new data set and the master data set, to make them compatible. Some of these changes are carried out in phase 7, and the new descriptor records are added to the master descriptor data set. Other changes are carried out in phase 8, and the new image data records are added to the master image data set. A full description of the logical operations within each phase is given below.

#### Phase 0

In phase 0, the input data from the digitizer and encoder are edited for validity and internal consistency. It should be noted that the design of this phase is under active development.

#### Phase 1

The phase 1 process operates on the raw map image data produced by the drum scanner. It uses the map control block (MCB) file to provide information about the map being processed (actual size of map, type of information carried, etc.).

The purpose of phase 1 is to eliminate redundant information in the raw map image data, and to identify and code the essential map elements (lines, areas, intersections, etc.) that need to be carried on for further processing. It also generates certain descriptive information about that map and passes it on to aid further processing.

In general terms, phase 1 accepts the map image data and automatically divides it up into small blocks, or "sections", that can be conveniently handled by the computer being used. Within each section it produces X and Y coordinates for all points lying along the lines of the map. It labels the line segments so described, and identifies and labels the points where they intersect other lines, segment boundaries, and map boundaries. It assigns an arbitrary number code (a "colour") to the areas enclosed by the lines within each section. It notes when areas in adjacent sections have been



given different numbers (colours) and produces a table of colours that are equivalent. It also identifies and labels the four corners of the map (Fig. A.2).

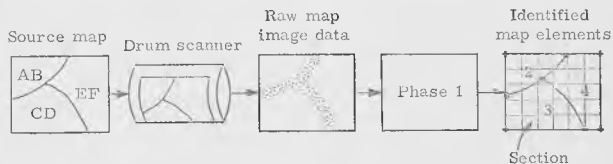
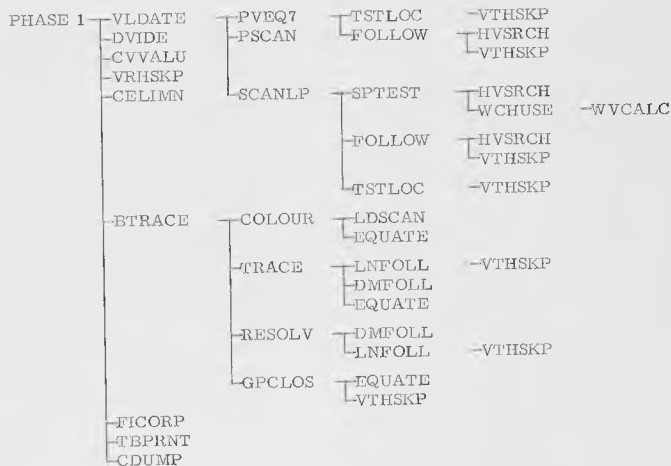


Fig. A.2. Phase 1 inputs and outputs.

The sequence of logical operations carried out in phase 1 is given in the flow chart overleaf (Fig. A.3).

Phase 1 is made up of a set of programs that work in concert to complete the sequence of logical operations. The names of the routines associated with each logical operation are indicated in Fig. A.3, and the hierarchical sequence for calling up the routines is listed below.



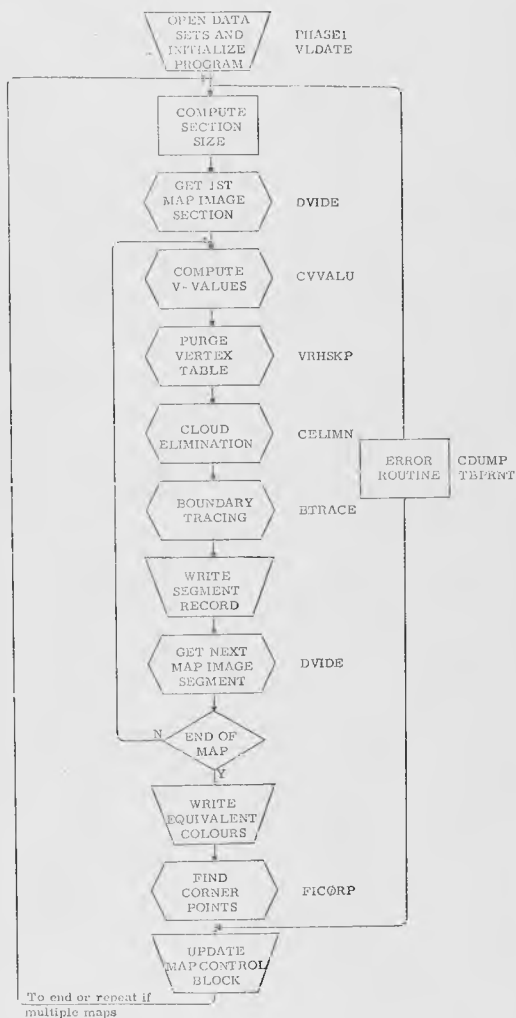


Fig. A. 3. Phase 1 flow chart.

The logical operations carried out in phase 1 are as follows:

Initialization and control. These functions are handled by the routine named PHASE1. This routine allocates core space to various tables, performs most input/output functions, and acts as a calling routine for the other groups of routines. A debugging facility is provided through routines CDUMP and TBPRNT. Routine VLDATE provides version and level numbers useful in keeping a record of program changes.

Compute section size and get first section. It is not possible to hold in core at one time all the points that make up a map, because a 20 in. x 30 in. map may comprise over 56 million bits of raw image data (over 7 million bytes of computer storage). The map must therefore be divided into smaller parts, called sections. Since a map is digitized by strips or columns, it might appear advantageous, in terms of input/output processing time, to handle the map on that basis. However, such an advantage would be lost when many small line segments were read across the resulting long, narrow sections. A square or nearly square section gives longer line segments and fewer connections between sections, so that it results in shorter processing time.

The program DVIDE automatically calculates the section size within limits set to use as much available core as possible. The overall size of the map is read from the map control block. A normal section size, utilizing about 350K of core, is 1.50 x 2.24 in. Sections are subsequently processed one at a time, starting at the lower left corner of the map and proceeding upwards and to the right.

Compute V-values. As described above, the raw map image data for each section is made up of an array of bits of data. The lines of the original map are represented by clouds of 1-bits, and the "background" of the map is filled in with 0-bits. The background information and the most of the clouds of 1-bits must be eliminated at an early stage so that lines, a single point in thickness, remain. Accordingly, a simple arithmetic operation is carried out by program CVVALU, which assigns a new value (termed a V-value) to each point in direct proportion to the number of 1-bits adjacent to it, as shown in Fig. A.4.

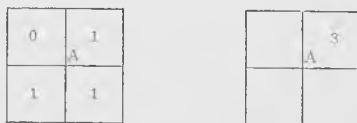


Fig. A.4. Computation of V-values; the V-value for point A is 3.

In this diagram, the V-value for a point is nominally written in the square with the given point at its lower left corner. The V-value normally ranges from 0 to 4, inclusive. The value may be artificially incremented during processing.

The V-value calculation has the effect of enhancing the centre points along the lines, as well as minimizing the effect of random-noise bits generated by dust particles on the scanned map. An example of this V-value allocation and the subsequent cloud elimination and boundary tracing process is given in a series of diagrams at the end of this appendix (Figs. A. 61, A. 62, A. 63, A. 64, and A. 65).

Vertex table initialization. To set the stage for cloud elimination, boundary tracing, and intersection identification, a table of line intersections or "vertices" must be initialized at this point, using routine VRHSKP.

Cloud elimination. During this part of the sequence of operations, each of the lines in the section is located and followed to a conclusion, guided by the V-values created earlier. Each line termination is entered as a vertex into a vertex table. These processes, controlled by routine CELIMN, are handled by a group of routines called PVEQ7, PSCAN, SCANLP, SPTEST, HVSCH, WCHUSE, WVCALC, FOLLOW, TSTLOC, and VTHSKP.

In this operation, an 8-bit-wide vertical scan procedure is initiated whenever it is necessary to start the cloud elimination of a new line in a section. A V-value of 2 is sufficient to stop the line-finding scan and commence "line following". The line-following process ensures that no point is selected as part of the line if there is an adjacent point that has a higher V-value. The procedure follows the line from the starting point and records (in the boundary trace image) the direction from one point on the

line to the next. This direction is coded as a "d" code (Fig. A.5).

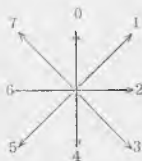


Fig. A.5. Direction of line-following scan ("d" code).

The line is followed until one of several things happens: the line being followed intersects itself, or one previously followed; it reaches a section limit; or it simply ends. In each case an identifying entry is made in the vertex table. Five types of vertices are recognized in phase 1:

- (1) E-vertex - a "dead end" vertex, with only one line entering or leaving it.
- (2) R-vertex - an arbitrary vertex assigned to a line that is a closed loop or "ring", so that one end of the line leaves the vertex and the other end of the same line enters it.
- (3) N-vertex - a "normal" vertex, which occurs at the intersection of two or more lines. It may have three to seven lines entering or leaving it. For convenience, two-line normal vertices are classed with R-vertices.
- (4) P-vertex - a vertex formed when a line crosses the outer "perimeter" of a section. It normally has two lines associated with it but may have up to seven.
- (5) C-vertex - a vertex that marks a "corner" of a section and normally has no lines associated with it, although it may have up to seven.

Having followed the line from the starting point to a logical vertex, the search process returns to the start and begins tracing in the opposite direction. However, to preserve the directional sense of the line and to keep the direction codes in the same convention for the whole line, the code placed in the trace record for the second half of the line (Fig. A.6) is the reverse of that used in the first half (Fig. A.5). The reverse code, it can be seen, is simply (d code + 4) modulo 8.

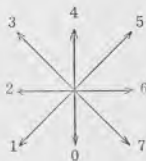


Fig. A. 6. Direction of second half of line-following scan ("d" code + 4).

When a line has been traced to a vertex at both ends, the line-finding scan for another starting point resumes from the previous one. Lines that have been followed and recorded are made "invisible" in the V-value section, so that a line cannot be traced twice. The method of scanning from start points ensures that all lines are followed. After an entire section has undergone cloud elimination, the boundary trace image will contain the centre points of all lines contained in the section and the vertex table will contain the location of all intersections and edge points.

Boundary tracing and colouring. Once the lines have been located, a number or "colour" is assigned to each area enclosed by the lines in the section. The lines are then retraced and the X and Y coordinates of the points are recorded. Some new vertices may be created in the process, and some small gaps in the line may be closed. These functions, controlled by routine BTRACE, are handled by the group of routines COLOUR, EQUATE, LDSCAN, TRACE, LNFOLL, DMFOLL, RSOLV, GPCLOS, and VTHSKP (also used by CELIMN).

The portion of a line lying between two vertices is known as a "segment". Every segment has a start (or opening) vertex and an end (or closing) vertex. A "lattice" is a set of vertices in a section that are all interconnected. The lattice that includes the P-vertices (see Cloud elimination above) is always the first lattice recorded in the section. (The perimeter lines of the section are classed as segments for the purposes of inter-connecting vertices in a lattice.) A "face" is the area bounded by a closed loop of segments. Each segment is associated with two faces, one on each side of it.

In "colouring", identifying tags are applied to both "sides" of the line segment. Sides (right or left) have meaning with respect to the direction

in which the line was first traced during cloud elimination. The identifying tags are called colours because they are analogous to the colours on a political map. In practice, the tags are attached to the vertices. A sort-and-search operation on these colours enables the segments to be connected and, hence, the faces to be assembled.

The lines are retraced and X and Y coordinates are assigned in a straightforward operation. The point of origin of the coordinate system is the lower left side of each section, and a unique number is assigned to each segment as it is traced. The number of points in each segment is recorded.

As mentioned above, the perimeter line of a section, that is, the common line between two adjacent sections, is known as the p-line. For continuity between sections, adjacent sections are overlapped to provide a common area eight points wide. The overlapped area straddles the p-line of the section with four points on either side, so that colours can be passed from section to section. This is carried out automatically between sections processed sequentially, but in some cases the sections do not come together and the match cannot be noted until several intermediate sections have been processed. D. Lever developed much of the logic involved in cloud elimination and boundary tracing.

Write segment record and get next segment. All pertinent information about the segment being processed is written in the boundary trace file, and the subroutine DVIDE calls up the next segment. This process is repeated until all segments in the map have been processed.

Write equivalent colours. Different colours noted as being assigned to each half of one face are listed in the "colour equivalence table", which is passed to phase 3 to be resolved.

Corner point location. The points at which the map borders intersect the section lines are kept in a working table. When all sections have been processed, four straight lines are fitted to these points and the intersections of the lines are taken as the corner points of the map. This is done by program FICORP. The corner points so determined are recorded in the MCB in step 12.

Error routine. Depending on the severity of an error, the map or the entire program may be terminated. Each error type has a unique code, which is printed at the end of the map processing and is also placed on the error file used by phases 1 to 3.

Update of map control block. The highest colour code, the last section number, the section size used, and the corner point locations of the map are added to the map control block, which is passed to phases 2 and 3 to provide parameters for further processing.

The outputs from phase 1 are thus:

- (1) A list containing all the coded line segment information
- (2) A list of equivalent colours between sections that are as yet unresolved.
- (3) An updated map control block.
- (4) An error file that accumulates the errors found in phases 1 to 3 for resolution in phase 4.

The following limitations apply to phase 1, as currently implemented. It is possible to rearrange some of these restrictions, depending on the availability of core storage.

Section size	Minimum 0.86 in. x 0.42 in. wide maximum 1.5 in. x 2.24 in. wide of frame height x frame width
No. of section rows or columns	maximum 64
No. of segments/section	maximum 250
No. of points/section	maximum 8000
No. of islands/section	maximum 59
No. of vertices passed to a section	80
No. of vertices on section boundaries for a column plus the top of one section	maximum 320

For the above restrictions, the following assumptions are made where required:

Section parameter	= 5
Table parameter	= 4
Map height	= 22 in.
Section size	= 1.50 x 2.24
Region	= 320K



In certain situations during line following, gaps that were not present on input appear in the output image. The situation, shown in Fig. A. 7, can be corrected only by use of the manual error correction (MEC) package.

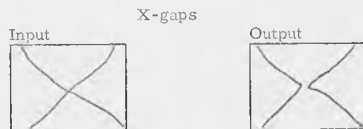


Fig. A. 7. Creation of gap in output image.

D. Lever, G. Morton, H. Lerch and A. Aus made significant contributions to the detailed design and programming of phase 1.

#### Phase 2

The coordinate system used in phase 1 to describe the image data line segments in each section is derived from the drum scanner, whereas the coordinate system used to describe the reference points related to each piece of descriptive data is derived from the table digitizer. These two coordinate systems must obviously be equated before the descriptor data can be matched to the image data and a data bank created, as data from both sources are involved. It is economical at this stage to convert both coordinate systems to the one used for main storage in the data bank. Phase 2 partially carries out this conversion, and prepares the way for its completion in phase 3.

The coordinate system used for the main storage of the data bank is the natural geographic coordinate system (GCS) of degrees of latitude and longitude. A particularly desirable feature of this convention is that it favours no one projection. Because it is neutral, transformations to and from the system for any projection involve only one step. More importantly, it is strictly conformal. There is no difficulty in going from one part to another, which is a significant advantage when a data bank covering a sub-continent is contemplated. When the earth is represented in either map projections or natural coordinate systems, the question arises whether it should be taken to be a sphere or an ellipsoid. This question was examined

and it was found that if 0.25% errors in both area and length calculations for areas up to 3,000 square miles are acceptable, then representing the earth as a sphere is adequate. These errors are virtually independent of the size of the quadrilaterals contemplated, and were considered acceptable for the purposes of the Canada Geographic Information System. The only remaining questions to be answered are ways of computing areas, lengths, and distances in the natural geographic coordinate system.

The difficulty encountered in performing area calculations in the GCS is due to the non-uniform grid length along the abscissa. Although this length is very nearly a cosine function of the latitude, over small changes in latitude, as a first approximation the relationship can be assumed to be linear. Considering two coordinate systems,  $S_1$  and  $S_2$  shown in Fig. A.8, where the X grid length of  $S_2$  is a linear function of the ordinate, it can be shown that a simple relationship exists between the area  $A_1$  in system  $S_1$  and the same area  $A_2$  in system  $S_2$ . If  $U_c$  is the abscissa length in system  $S_2$  at the centroid of area  $A_1$  in system  $S_1$ , then  $A^2 = \frac{U_c}{X} \times A_1$ . No additional effort is required to calculate  $U_c$  on an ellipsoid as opposed to a sphere. Since any large area can be subdivided into smaller areas, where the first approximation is virtually exact we can calculate any smaller area to any degree of accuracy. There is a small error due to the change in the Y grid length because of the ellipticity of the earth, but this effect is several orders of magnitude smaller than the effect of the X axis. However, to complete the procedure for calculating areas, it should be noted that the Y grid length at the centroid can be used as the average Y grid length.

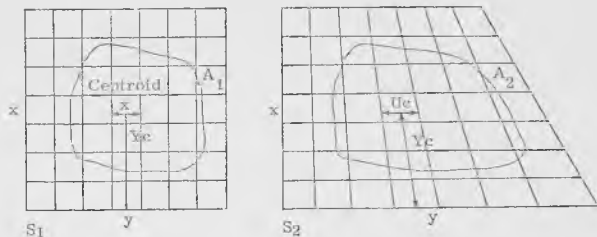


Fig. A.8. Relationship of areas in two coordinate systems.

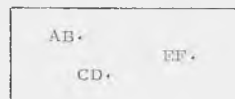
To see how closely the above formula agreed with the true value of area on an ellipsoid, the percentage error in area was tabulated as a function of the angle of latitude subtended by the area. To obtain an estimate of the best and worst cases, two shapes of figures were considered, one a square and the other an L shape one grid element in thickness. A grid length of  $1/16384^{\circ}$  was taken, which represents a distance of about 22 feet in the meridian and 11 feet in the parallel of  $60^{\circ}$  latitude. It was observed that, for areas with a range of 500 miles in the meridian and 300 miles in the parallel at  $60^{\circ}$  latitude, the maximum error incurred was less than 25 parts in 100,000.

Distance calculations in the GCS cause no serious difficulties. If regions are sufficiently small the Pythagorean theory can be used, whereas for greater distances a great-circle approximation can be employed for an equivalent sphere whose radius is the mean radius of curvature at a latitude midway between the two points. Still more accurate determination of distances can be obtained using standard but more complex procedures. The latter are not utilized in CGIS at this stage of development. F. Jankaluk was responsible for the mathematics and programming underlying the use of GCS.

Input to phase 2 consists of the segment records contained on the boundary trace file, output from phase 1, and the edited descriptor data and related reference coordinates output from phase 0. In addition, the map control block provides identification information about the map being processed.

The output from phase 2 is, in effect, a copy of the descriptor data with their reference coordinates within each section changed to the GCS format, together with the transformation coefficients to be used in phase 3 to complete the conversion of the image data to GCS format. The map control block is also updated by phase 2 (Fig. A.9).

Edited descriptor data  
with related reference  
coordinates



Descriptor data with  
reference coordinates  
in GCS, by section



Fig. A.9. Conceptual functions of phase 2.

The sequence of logical operations carried out in phase 2 is illustrated by the flow chart in Fig. A. 10.

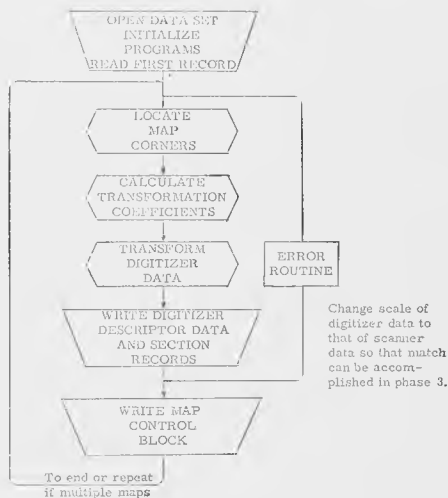


Fig. A. 10. Phase 2 flow chart.

Phase 2 consists of a main program (PHASE2) that controls the execution of the routines and subroutines, and handles much of the computation and all of the input/output. There are four main routines in phase 2 (RADUTM, UTMRAD, XCOEFF, and ERRANL), which are used for repetitive calculations and error handling.

Phase 2 performs the following logical operations:

Initialization and control. These processes are handled by the main program, PHASE2, which allocates and releases core as required, performs all input/output functions, acts as a calling routine for the subroutines, and performs computations not directly related to transformations and error analysis.

Location of map corners. In order to perform any transformations involving the scanned map data, it is necessary to locate the map corners accurately. The process involves reading the map corner points from the digitizer data (taking the mean of several readings), taking the map corner points as determined from the image data in phase 1, taking the latitude/longitude coordinates of the map corners recorded in the map control block, and relating all these. This is done in PHASE2, utilizing routine RADUTM where necessary to relate the actual map to the corresponding theoretical UTM map.

Compute transformation of coefficients. Two sets of transformation coefficients are computed in phase 2:

- (a) Digitizer and scanner transformation coefficients
  - required to transform descriptor-related reference coordinates into equivalent scanner coordinates.
- (b) Distortion/orientation transformation coefficients
  - required to compensate for any errors introduced by distortion of the map sheet, or by the orientation of the actual map on the scanner.

Both transformations are co-linear transformations, and the coefficients are computed in routine XCOEFF.

Transform descriptor-related reference coordinates to GCS. Using the already calculated transformation coefficients, each of the descriptor-related reference coordinates is transformed into scanner coordinates to determine into which section of the reference coordinate points fall. The points are then transformed for distortion and orientation and finally into GCS coordinates. The points are passed to phase 3 in this form.

Write out revised descriptor data records and section records. Two records of the revised descriptor data are generated for use in subsequent processing. Each record contains the transformed reference coordinates and the section number in which it falls. In addition, one of the records (DS3507) contains the face number and full descriptor passed from the source data. When all faces have been processed, one record is written on the boundary trace file for each section of the map. These records contain the section number, the count of reference coordinates in the section, and the transformation coefficient array.

Error analysis. When an error is detected during the run, an error number (ERRNUM) is assigned and a call made to subroutine ERRANL. The subroutine assembles an error message based on the error number and writes it on file DSERROR. Depending on the severity of the error, the remainder of the input records for the current map may be bypassed.

When control returns from ERRANL, the program will continue, terminate the map, or terminate the run, depending on the type of error.

Write map control block and end of phase. When the map has been completely processed (or terminated because of an error), the error code, run date, and version and level numbers are added to the MCB and the updated record is rewritten. All "allocated" core is released, and the error code is reset.

The outputs from phase 2 are thus:

- (1) A list of descriptor data in each section with the related reference coordinates in GCS. Before being passed to phase 3, this list is sorted into order of sequentially numbered reference coordinates within each sequentially numbered section within each map.
- (2) A list of reference coordinates in each section. Before being passed to phase 3, this list is sorted into the order of sequentially numbered reference coordinates within each sequentially numbered section within each map.
- (3) A list containing all the coded line segment information in each section to which has been added the count of reference coordinates in each section and the coefficients necessary to transform the X-Y coordinates of the line segments to GCS.
- (4) The error file updated with errors unresolved in phases 1 and 2.
- (5) An updated map control block.

Note: The phase 2 algorithm for locating corner points is accurate to  $\pm 1$  scanner point for clean map corners. The accuracy is reduced if any of the following situations occurs:

- (a) Dirt or spurious lines close to the map corner are present.
- (b) The map borders are not drawn in the projection being used.

- (c) In the given projection, no part of the border from a corner to a point on the border approximates a straight line.

In the present implementation (UTM), the first 5% of any border must lie along a point "band width" of 2, at a resolution of 250 points to the inch.

Transformations. Within each section, only the section corner points are accurately transformed from scanner to GCS format, and the remaining points in the section are extrapolated linearly from them. This causes negligible distortion for sections within the size range allowed in phase 1.

B. Kemeny, I. Grossman, and A. Aus made significant contributions to the detailed design and programming of phase 2.

### Phase 3

Phase 3 receives as input a data set that contains the "image" of the original map and a data set that contains the descriptor data of that map, together with a single point for each descriptor, which has now been oriented to the image data. Phase 3 links the two data sets and so conceptually restores the map to its original classified status. It also changes the basic unit of data storage from the "section" that was convenient in phase 1 to the "frame", a larger unit convenient for the final data bank.

These concepts can be expressed graphically, as shown in Fig. A. 11.

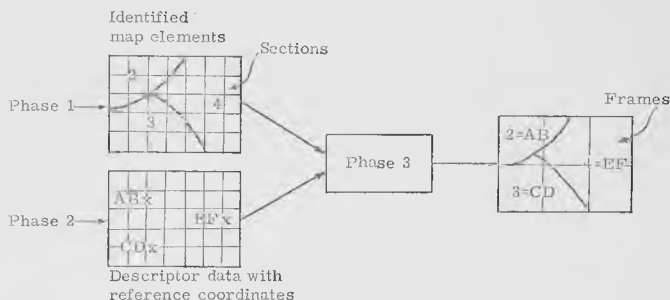


Fig. A. 11. Inputs and outputs of phase 3.

The sequence of logical operations carried out in phase 3 is illustrated in the following flow chart (Fig. A. 12).



Fig. A. 12. Phase 3 flow chart.



Phase 3 is made up of a set of programs that work in concert to complete the sequence of logical operations. Specifically, it consists of five routines, external to each other, compiled separately and link-edited together. They are known as follows:

<u>Source program name</u>	<u>External entry name</u>	<u>Invoked by</u>	<u>Invoke</u>
AB1	PHASE3	(Options MAIN)	AB2, AB5
AB2	PH3SEC	AB1	AB3, AB4, AB5
AB3	PH3SEG	AB2	
AB4	PH3FRM	AB2	
AB5	PH3TRC	AB1, AB2	

The following diagram attempts to depict the calling-sequence relationship of the five routines (Fig. A. 13).

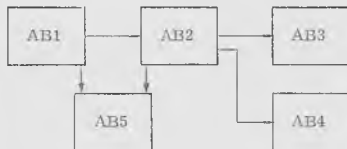


Fig. A. 13. Calling sequence of routines in phase 3.

The routines used in each logical operation are summarized below.

<u>Logical operation</u>	<u>Routine</u>				
	<u>AB1</u>	<u>AB2</u>	<u>AB3</u>	<u>AB4</u>	<u>AB5</u>
Framing	x	x	x	x	
Colour equivalence	x	x			
Linkage	x	x	x		
Transform	x	x	x		
Scale reduction			x		
Gap closing			x		
Smoothing			x		
Compacting			x		
Geodetic calcs.			x		
Tracing	x	x	x	x	x

The programs are designed to be efficient in terms of the computer architecture, and to take advantage of the operating system. For the purposes of this brief technical description of the overall system, the contribution of phase 3 can best be understood from an examination of the logical operations.

Initialization and control. These functions are handled by the AB1 routine, which allocates core to the various tables, performs most input/output functions, and acts as a calling routing for the other groups of routines (see above).

Framing. To understand the structures inherent in framing, it is necessary to examine the concepts of scale factor, frame factor, and edge factor, as used in the system.

The smallest division in the geographic coordinate system is a "unit grid"; its size is derived quite empirically. The greatest extent of the Geographic Coordinate System needed to encompass Canada is  $128^{\circ}$  in longitude. A full word of storage in fixed binary notation contains 32 bits, so that eight bits of binary notation can represent the  $128^{\circ}$  and the remaining 24 bits represent the possible subdivisions of any one degree. The finest subdivision in the system (the unit grid) is thus an angular distance of  $2^{-24}$  degrees in longitude and latitude. At  $45^{\circ}$  latitude this is approximately 1/6 in. longitudinally (X direction) and 1/4 in. latitudinally (Y direction). In practice, a subdivision coarser than the unit grid is desirable. The angular distance between two points actually stored is thus a multiple of the unit grid, that is, some multiplication between 0 and 31. The degree of multiplication is called the scale (S) factor. The optimum S factor obviously depends upon the degree of resolution of the source data, and on the degree to which it is desired to maintain or decrease that resolution in system storage. An S factor of 9 (one point every 0.1046 sec, or 133.5 in. x 86.2 in.) is used in the CGIS for data derived from 1:50,000 maps.

It will be remembered that in phase 1, the raw image data was divided up into the largest units convenient to process. These are the relatively small "sections" in which the files are still organized at the beginning of phase 3. Because much redundant information was eliminated in phase 1, it is convenient and efficient at this stage to group the remaining data into larger units. Such units, larger than sections but smaller than a map, are called "frames", and they are the main unit of storage for the final data bank. Their eventual size depends on the resolution of the data to be stored, and the capacity of the computer central processing unit. It is

also efficient if such frames fit into the structure of the Geographic Coordinate System. A square was found to be the area most efficiently shaped, in terms of maximizing the line segment lengths contained within it and ease of subsequent calculations and grouping. A unit frame is thus a square (in terms of the GCS) having an arbitrary maximum of  $2^{16}$  unit grid points along each side (14.1 sec, or 1,423 ft. lat. x 919 ft. long. at  $50^{\circ}\text{N}$ ), as shown in the diagram (Fig. A. 14). The unit frame of  $2^{16}$  was chosen as being the maximum number that can be represented in a one-half-byte word. In practice, a grouping larger than the unit frame is desirable and the actual frame used is a multiple of the unit frame. The degree of multiplication, called a frame (F) factor, can have values in the range of 0 to 16, inclusive. The upper restriction is controlled only by the fact that  $2^{16}$  is the largest number that can be represented in a half-word fixed binary field. An F factor of 5 is used in the CGIS for data derived from 1:50,000 maps. (A factor 5 frame is  $7.5 \text{ min}^2$ , or approximately 3.6 miles lat. x 5.6 miles long. at  $50^{\circ}\text{N}$ .)

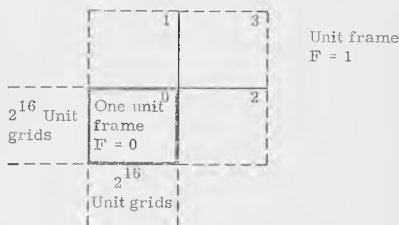


Fig. A. 14. Relationship of unit grids to unit frames.

The actual number of grid points along the side of an actual frame obviously depends upon the choice of S and F factors. This edge (E) factor can be expressed as:

$$2^E = 2^{16} \text{ when } S = F = 0$$

The relationship between E, S, and F factors can thus be expressed as:

$$E = 16 - S + F$$

E is limited to values of 0 to 15, so that the X and Y coordinates within a frame may be represented in 16 bits. A practical limit of 11 exists if a

core bit image is to be laid out in 512K bytes of storage.

Each unit frame in the system is assigned a unique number starting at the origin (lat.  $145^{\circ}\text{W}$  and long.  $40^{\circ}\text{N}$ ). From that point, frames are sequenced so that they fan out from the origin as frame number increases. The arrangement is shown in Fig A. 15.

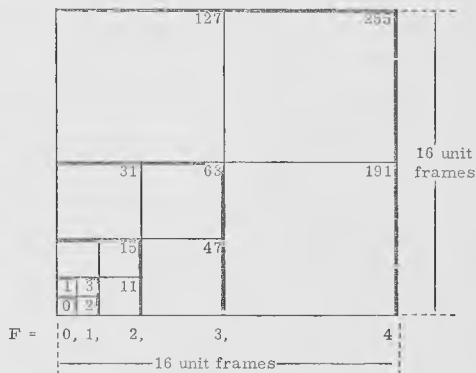


Fig. A. 15. Sequence of frame numbers.

The diagram (Fig. A. 16) shows the arrangement of the first 256 unit frames. This scheme, named the Morton matrix after its designer, G. Morton, has the advantage that frames close together in the sequence will probably be close together on the earth's surface. Equally, records close together on earth have the minimum separation possible in a sequential file. This can materially reduce search time, especially when small areas are involved. An additional benefit of this file structure is that the actual record addresses of a geographical location can be directly computed by manipulating the binary representation of the geographic X-Y coordinates shown on the Morton matrix diagram. The address is established by interleaving the binary representation of the X and Y coordinates in such a way as to combine the individual digits of each axis in sequence from a high order to a low order.

	85	87	93	95	117	119	125	127	213	215	221	223	245	247	253	255
	84	86	92	94	116	118	124	126	212	214	220	222	244	246	252	254
	81	83	89	91	113	115	121	123	209	211	217	219	241	243	249	251
	80	82	88	90	112	114	120	122	208	210	216	218	240	242	248	250
	69	71	77	79	101	103	109	111	197	199	205	207	229	231	237	239
	68	70	76	78	100	102	108	110	196	198	204	206	228	230	236	238
	65	67	73	75	97	99	105	107	193	195	201	203	225	227	233	235
	64	66	72	74	96	98	104	106	192	194	200	202	224	226	232	234
111	21	23	29	31	53	55	61	63	149	151	157	159	181	183	189	191
110	20	22	28	30	52	54	60	62	148	150	156	158	180	182	188	190
↑ 101	17	19	25	27	49	51	57	59	145	147	153	155	177	179	185	187
Y 100	16	18	24	26	48	50	56	58	144	146	152	154	176	178	184	186
3 011	5	7	13	15	37	39	45	47	133	135	141	143	165	167	173	175
	0101	0111	1101	1111												
2 010	4	6	12	14	36	38	44	46	132	134	140	142	164	166	172	174
	0100	0110	1100	1110												
1 001	1	3	9	11	33	35	41	43	129	131	137	139	161	163	169	171
	0001	0011	1001	1011												
0 000	0	2	8	10	32	34	40	42	128	130	136	138	160	162	168	170
	0000	0010	1000	1010												
	000	001	010	011	100	101	110	111								
	0	1	2	3	X	→										

Fig. A. 16. Morton matrix.

In the "framing" operations in phase 3, the first stage of conversion to a frame format is carried out by routines AB1, AB2, AB3, and AB4. All P-lines (section perimeter lines) are discarded and a new type of artificial line called frame lines (the frame borders) are generated. (Map borders will eventually have to be discarded and they will be replaced by frame lines, as one of the constraints of the system is that all map borders fall on frame borders.) Phase 3 eliminates all P-lines but does not eliminate vertices that have been caused by intersections with section lines; that step is carried out in phase 4. The phase 3 operation is illustrated in the diagram below (Fig. A. 17).

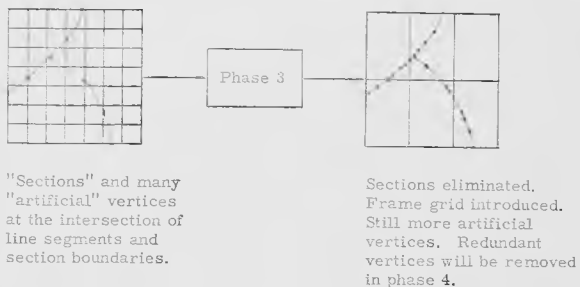


Fig. A. 17. Phase 3 operations.

Colour equivalence resolution. As the sections are brought together into frames, the colour equivalence table generated in phase 1 is used to resolve most of the duplicate colours assigned in that phase. Frames are substantially larger than segments, and two types of colour equivalence resolution are carried out (Fig. A. 18).

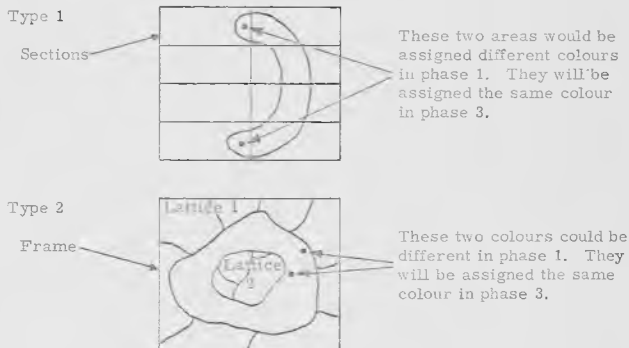
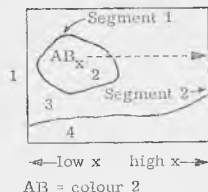


Fig. A. 18. Colour equivalence resolution.

The colour with the lowest numerical value in each set of equivalent colours is the one to which the equivalence will be resolved. Routines AB1 and AB2 are involved in this operation.

Linkage. This is the basic process of linking the image data from phase 1 to the descriptor data from phase 2, and conceptually recreating the source map in computer storage. The task of phase 3 is to discover into which face in the image file each of the descriptor-data reference coordinates falls, thus linking each descriptor to a face colour. The algorithm assumes that the data structure is arranged so that all areas are defined by sets of segments in a table, and that the entry for each segment identifies an arbitrary "start" vertex and an "end" vertex and contains an identifier to indicate on which side of the segment the area is found.

The process of determining if a point is within an area is started by proceeding along a line from the point, parallel to the axis, point by point, until an intersection with a boundary segment is met. The search then proceeds around that intersection until the point containing the next boundary (in any direction) is found. The process is repeated and the boundary followed until the next normal vertex is found. The coordinates of this vertex are stored. The same line segment is then retraced and the coordinates of the other vertex are determined. The two vertices are then compared with the vertices of the line segments in the table, and when a match is made, the line segment is identified. From the ordering of the vertices on the table into start and end vertices, the side of the line on which the point lies can be determined. The boundary of the entire area is then followed around in the same fashion, to ensure that the point is contained within the area concerned (Fig. A. 19).



A probe runs from the reference coordinate AB to the edge of the map. It encounters segment 1 and segment 2, but takes its colour from segment 1 because segment 1 is nearest to the reference coordinate. Since segment 1 carries two colours (2 and 3), phase 3 must decide which colour falls on the "low-x" side of segment 1.

Fig. A. 19. Linkage process.

Gap closing. Because the change of resolution required is usually downward, several points of input, after transformation and scale reduction, will fall on to a single point of output. As a result of rounding in the transformation routines, a gap will sometimes be created in the output segment where none existed in the input. It is a task of phase 3 to close such gaps with straight lines, and it is carried out in routine AB3 (Fig. A. 21).

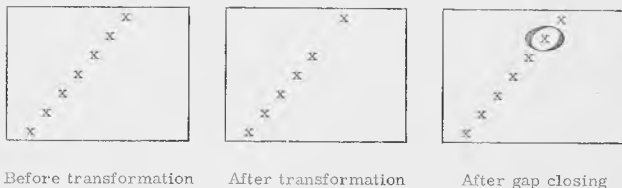
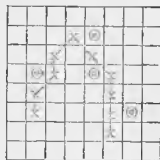


Fig. A. 21. Gap closing.

Single point smoothing. After transformation and resolution change, a line often has "jagged points", which can be defined as points in a line whose neighbours on either side are adjacent, and therefore could be joined without passing through the jagged point (Fig. A. 22).



In the diagram, the "jagged" points are circled. They could be removed, and the line shown will still have no gaps in it.

Fig. A. 22. Single point smoothing.

Since, at a resolution of 250 points per inch, these jagged points could not have been drawn intentionally, it is assumed that they are accidental. They would add heavily to the amount of storage required to keep the compacted representation of the line, and they would add to the timing required to compact and uncompact them; it is therefore convenient to drop such points. The process of dropping these points is called "smoothing".



Transformations. Using the transformation coefficients determined in phase 2, the image transformations are completed in phase 3. Specifically, in phase 3 two types of linear distortion in the image are eliminated and the corrected result is converted to GCS (Fig. A. 20).

Linear distortion type 1



Linear distortion type 2

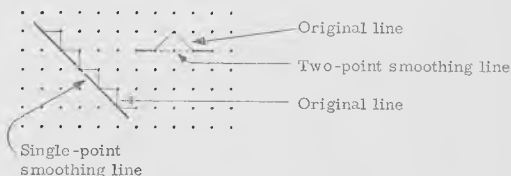


Fig. A. 20. Transformations.

The algorithms required to perform these two operations are combined into a single algorithm, which is carried out by routines AB1, AB2, and AB3. However, because applying that algorithm to every point within a map would be an extremely expensive use of computer time, a compromise is reached by applying the actual algorithm to the corners of each section only, and calculating points within the section as linear interpolations from corner points. The inaccuracy incurred is not significant in system terms.

Resolution reduction - scale change. The maps input to the system have been scanned at a resolution of exactly 250 points per inch of map. The data bank resolution may vary, but will seldom be more dense than 50 points per linear inch of map. An additional task of phase 3 is therefore to reduce the resolution of the incoming map; the process is combined with the transformation functions into the single algorithm. Change of scale (i.e., the actual distance portrayed per inch of map) may also be required, and this is accomplished by the transformation routine AB3.

The difference between single-point smoothing and more sophisticated smoothing, or generalization, is illustrated in the following diagram (Fig. A. 23).



Note: A single-point smoothing line cannot pass through a point other than those on the original line.

Fig. A. 23. Two-point smoothing or generalization.

Single-point smoothing is carried out in concert with the other transformations in routine AB3.

Compacting. In the line segment file passed from phase 1, the line segments are described as a series of coordinates of each of the points comprising the segment. For a normal section size, a full word is needed to describe each point in the segment. One average 20 in. x 20 in. map containing 800 inches of line segment will occupy approximately 200,000 bytes of storage if this notation is employed. When it is considered that not one map but several thousand are entered into the system, a more efficient form of storage must be utilized. R. L. Kemeny devised the compact notation used in CGIS. The following brief comments will describe the process as carried out in routine AB3 in phase 3.

The technique employs sequences of two-bit codes denoting, by context, either a change of direction or a number of steps in an established direction. Four direction changes or three units of distance can be represented in two bits, and because a distance of zero has no meaning, it can be utilized to give a fifth direction: no direction change.

The choice of which direction changes out of the possible eight will be given a two-bit code is based on the frequency of occurrence. It was observed that scribed lines on land use maps have a tendency to be smooth, and hence,

sharp backward turns of direction are rare. By allocating the four codes to direction changes  $-2(+6)$ ,  $-1(+7)$ ,  $+1$ ,  $+2$ , virtually all direction changes that will normally occur, except 0, have been accounted for. To allow for increased capability to compact straight lines, 0 has been given a special significance. Using the definition of d-code demonstrated below, the following basis for the compact notation can be formulated:

A fundamental unit of compact notation, except for special conditions of change in direction, is a four-bit, one-half byte. The d-code is defined to the  $k^{\text{th}}$  unit of compact notation to be  $D_k$  ( $D_0 = 0$  by definition). The direction change  $\Delta$  is further defined between the direction of the  $k^{\text{th}}$  unit and the  $k+1^{\text{th}}$  unit, taken in clockwise direction to be a positive number defined by the formula:

$$\Delta_{k+1} = (D_{k+1} - D_k + 8) \bmod 8$$

Also, we have:

$$D_{k+1} = (D_k + \Delta_{k+1}) \bmod 8$$

Using these definitions, the following list can be constructed:

Compact notation unit	$\Delta$	Distance
00XX	0	$(XX) \cdot 4^{(N-1)}$ units
XX01		XX units
XX10		XX units
0000 0110 XX00	3 $(4+7)_8$	XX units
0000 0110 00XX	4 $(4+0)_8$	XX units
0000 0110 XX01	5 $(5+1)_8$	XX units
XX10	6	XX units
XX00	7	XX units

$N$  = Nth consecutive occurrence of  $\Delta = 0$   
 $XX = 01, 10, 11$

According to the above definition, the following would describe the line in the figure below (Fig. A.24).

k	$\Delta_k$	$D_k$	Compact	Result
1	1	1	0101	1 unit in direction 1
2	1	2	1101	3 units in direction 2
3	0	2	0010	$2 \times 4^0 = 2$ units in direction 2
4	0	2	0010	$2 \times 4^1 = 8$ units in direction 2
5	1	3	1001	2 units in direction 3
6	4	7	0000 0110 0001	$1 \times 4^0 = 1$ unit in direction 7
7	0	7	0001	$1 \times 4^1 = 4$ units in direction 7
8	7	6	1000	2 units in direction 6
9	6	4	1010	2 units in direction 4

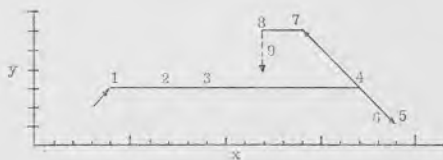


Fig. A.24. Compacting.

The list of X-Y's for this line requires 864 bits whereas in compact notation it occupies 76 bits, of which 32 are taken up with the start point X-Y coordinates. If required, lines exhibiting regular patterns can be further compacted by storing the pattern and the repetition factor.

Geodetic calculations. The data bank must provide information concerning each face of a coverage, such as area and centroid. Phase 3 calculates values for each segment (moment, measure); in phase 4 these values are accumulated or averaged for all of the segments comprising a face, thus yielding area and centroid results.

The geodetic calculations are carried out by routine AB3. The actual sequence of events is interleaved with the compacting process: (a) pre-compacting, (b) geodetic calculations, (c) compacting.

The area calculation is carried out by first determining the distance between the initial vertex point and the next vertex point on the perimeter, and then calculating the area subtended by the two points to the axis of the coordinate system. The area so calculated is added to the total of areas calculated in this manner around the entire perimeter, either in a positive or negative manner, depending on the direction of the segment. The direction of a segment is determined in relation to the axis subtended. If the difference between the first vertex of the segment examined and the second vertex of the segment examined is positive on the axis, the direction of the segment is considered to be positive. A negative change on the axis allocates a negative direction to the segment. The area calculation process can be made clear by reference to Fig. A.25. Starting at vertex 1 to measure the size of area A, the distance between vertex 1 and vertex 2 is

determined and the size of the subtended area  $P(1, 2, 2', 1')$  is calculated. The area subtended by all line segments is calculated in the same way. Between vertex 1 and vertex 5, all segments are positive with relation to the X axis. Segment 5, 6 subtends area  $N(5, 5', 6', 6)$ . As segment 5, 6 is negative in direction with relation to the X axis, area  $N$  is added in a negative manner to the total, that is, it is subtracted. The process is repeated for the area subtended by segment 6, 1 and the resulting calculation provides the area of  $A$ .

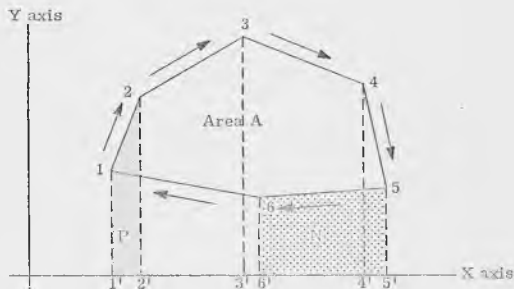


Fig. A.25. Level 3 measurement area calculation.

To determine the centroid, the total moment must be divided by the total area. The steps in the centroid allocation algorithm are as follows:

The subtended area of each increment along the axis is calculated,

$$\sum_{i=1}^n \sum_{j=1}^n (X_i - X_{i-1})(Y_j - Y_{j-1}) \text{ where } X_n = X_0 \text{ and } Y_n = Y_0,$$

as is the subtended moment using the area calculation itself,

$$\sum_{i=1}^n \sum_{j=1}^n \left( \frac{X_i + X_{i-1}}{2} \right) (X_i - X_{i-1})(Y_j - Y_{j-1}) \text{ where } X_n = X_0 \text{ and } Y_n = Y_0.$$

As each subtended area and subtended moment are calculated, they are separately summed. The process is repeated until the figure is completed. The total of the subtended moments is divided by the total of the subtended

area,

$$\bar{X} = \frac{\sum_{i=1}^n \sum_{j=1}^n \left( \frac{X_i + X_{i-1}}{2} \right) (X_i - X_{i-1}) (Y_j - Y_{j-1})}{\sum_{i=1}^n \sum_{j=1}^n (X_i - X_{i-1}) (Y_j - Y_{j-1})}$$

This provides the coordinate of the first axis. The process is repeated to determine the coordinate on the second axis,

$$\bar{Y} = \frac{\sum_{i=1}^n \sum_{j=1}^n \left( \frac{Y_j + Y_{j-1}}{2} \right) (X_i - X_{i-1}) (Y_j - Y_{j-1})}{\sum_{i=1}^n \sum_{j=1}^n (X_i - X_{i-1}) (Y_j - Y_{j-1})}$$

Error tracing. Because phase 3 represents the moment in the data reduction system at which the image and classification data are both available for the first time to one program, this phase is the logical point at which to investigate apparent errors. The tracing processes of phase 3 are designed to show whether errors have arisen from image or classification input data, whether phase 1 or phase 2 processes have caused the errors, or whether the data passed on to phase 4 are, in fact, correct. The tracing procedures also serve as a permanent method of debugging phase 3 errors. They are incorporated in all routines, AB1, AB2, AB3, AB4, and AB5, used in phase 3. Nevertheless, the phase 3 operation itself does very little to correct the errors.

The phase 3 operation checks the sequence of input and the record type and it does not accept maps for processing that are not accompanied by a map control block. It does not process any map whose map control block shows a high level of error in phases 0 to 2. It expects to find input records for all input files (except trace input), and will not process a map unless all files are present. In the case of trace input, if it is dissatisfied with the one or more trace cards, it will offer the operator the choice of termination or continuation without trace.

The system used by phase 3 to indicate the nature of its errors utilizes a six-character code, R3xxpe, where xx represents the type of error, p represents the external procedure within which the error occurred, and e represents the error-number within that procedure.

The outputs from phase 3 are thus:

- (1) An unsorted, linked file with the various types of records generated from several different places in phase 3. The coverage header, which is the last record written, carries data concerning the number of faces and descriptors processed. The frame records carry data concerning the number of faces and the segments encountered in each frame, the position of the frame in the output map, and the frames adjacent to it. Face records are written for each face encountered, once for each frame in which each face falls. Face records carry the descriptor of the face and the colour that has been linked to it. Two segment records are written for each segment processed or generated by phase 3, and carry all data concerning the segment, including the compact notation of the actual points comprising that segment.

This file is sorted into numerical sequence of colours within numerical sequence of frames, before being passed to phase 4.

- (2) An error file updated with the errors unsolved in phases 0 to 3.
- (3) An updated map control block.
- (4) Printed output on phase events, time start and end, end messages, etc., as requested.

The following basic programming limitations apply to phase 3:

- (a) Section width must be less than frame width.
- (b) Section height must be less than frame height.
- (c) All significant tables in phase 3 are allocated, based on data density. It is, therefore, difficult to define core requirements. However, there is a data density that will overflow any region size.

The following inherent distortions in phase 3 should be noted:

- (i) Aligned section and frame borders. Segments that did not end on a section border before transformation will be distorted by  $\pm 1$  point, if transformation causes them to end on a section border that is also a frame border.

- (ii) Alignment of map borders. Map information falling outside the outside GCS frame borders of a map is ignored. Maps will be stretched a maximum of  $n$  scanner points if the map border falls inside the outside GCS frame border. The phase 4 input card determines  $n$ .

A. Benjamin made significant contributions to the design and programming of phase 3, in particular the map error recognition and correction and the programming of the linkages between image and descriptor data.

#### Phase 4A

The main task accomplished by phase 4A is to build the records of the map produced in phase 3 into the general structure of the main data bank. As described in the introduction above, the data bank consists of two files: the image data set (IDS) and the descriptor data set (DDS). The common unit of storage in both files is the face. The IDS contains the image records of the face, that is, the line segments records that belong to each face, and the point records that make up each line segment. For each type of data within the IDS, the face records are listed within each frame as shown in Fig. A. 26.

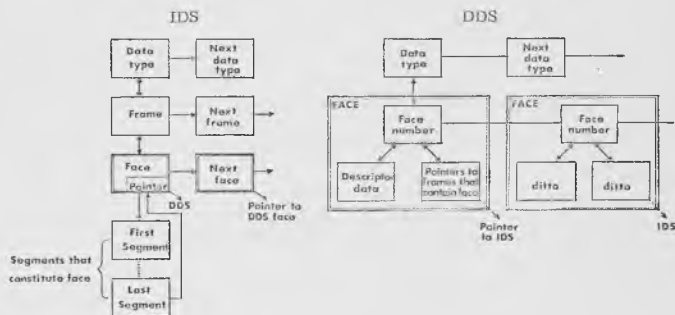


Fig. A. 26. Structure of IDS and DDS.

The DDS contains the descriptive records of each face for each data type. Linkages between the files are established by pointers. Each face record in the IDS contains a pointer to the part of the DDS that contains the



descriptive record for that face, and each face descriptor record in the DDS contains a pointer to the frame(s) and face in the IDS that contain the image data for that particular descriptor. (More than one frame may contain the partial records of the image data needed to describe the boundary of one face description contained in the DDS.) The function of phase 4A is thus to arrange these records and linkages. To carry this out, each face has to be clearly delimited and its component parts assembled. In the process, illogical conditions become apparent and are either corrected by procedures within the phase or are written on an error record. The error record produced by phase 4A is processed by the manual error correction (MEC) subsystem and the output from that process is recycled through phase 4A. Phase 4A is the gateway to the data bank, and this step ensures that data from individual maps are topologically correct before they pass to the sequence of operations that will insert them in the final data bank.

The flow chart (Fig. A. 27) gives an overview of the sequence of logical operations involved.

Phase 4A performs the following logical operations. As in other phases, the code needed to accomplish these tasks is to be found spread out throughout the phase.

Coverage initialization. This process is handled by the main program named PHASE4. Only one data type (coverage) may be processed in any one run, so there is only one coverage record to be read in, and all coverage-dependent fields are initialized on the basis of this record.

Initialize and process frame records. The first part of the frame record to be read in is the frame "header" (Type 1 frame record). There is only one of these per frame, and it contains information concerning the number of faces and segments in the frame. On the basis of the map number to which the frame belongs, the map control block is read in and the frame parameters are calculated from information found therein. The remaining frame records are then read in. These are type 2 frame records, and there may be up to four of them, one for each of the four frame sides which is a map border. Switches are set to indicate which of the sides are map borders, and these switches are referred to as needed throughout the remainder of the phase.

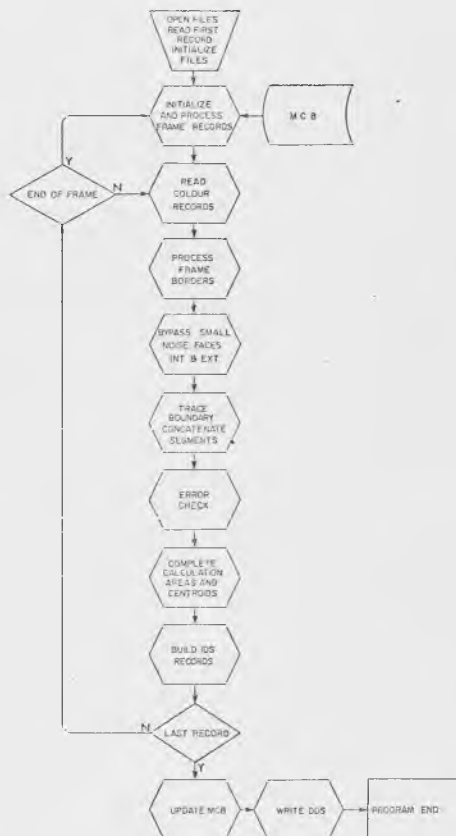


Fig. A. 27. Phase 4A flow chart.

Read colour records. The contents of the frame are processed next. The colour records, including their related segment and vertex records, are read in, one colour at a time, until the next frame record is reached.

Because it is the dividing line between two faces, each segment has two colours associated with it. The "major" colour is the colour of the face to the right of the segment, the "minor" colour is the colour of the face to the left.

Since the input file is sorted in ascending colour sequence, the first colour to be read in is colour 0. This is the colour assigned to the area that falls to the right of segments making up the frame border. Colour 0 represents all the area outside the frame, and can be regarded as a pseudocolour representing the frame border.

Colour 1 is a special case of colour 0, and is assigned to an area that falls inside the frame border but outside the map border, that is, an area exterior to the map that has fallen inside the frame border.

The third special colour is colour x, actually colour = 32,705 (COLORX). This is assigned in phase 3 to frame-border segments when the actual colour cannot be determined. These first three types of colour are handled in the logical operation "processing the frame border" before the normal colours are processed.

Processing the frame border. This logical operation is partly dependent on the degree of frame border error acceptable to the user of the system. CGSTOL is a variable whose value is determined by calculation, in phase 4, using the map parameters and the value assigned to SCNTOL. The value of SCNTOL is set at the time of compilation, and represents the number of points of tolerance on the scanner grid allowed to colour 1 inside the frame border before a map border gap is declared.

In the case of colour 1 segments, if the discrepancy between the map border and the frame border is less than CGSTOL grid points, then new segments are generated by projecting the old segments on to the frame border and aligning them with it. The new vertices are the old vertices projected on the frame border. In other words, the map border is forced to fit the

regular frame borders and the inaccuracy incurred (never more than three or four grid points) is accepted. This process may create minor gaps in the border, which will be filled either by other existing segments or by segments generated by means of routine NSLOC3.

If the discrepancy between map border and frame border is more than GCSTOL, then a major map border gap is inferred and all the segments concerned are listed in the error output file.

Colour x segments are replaced by segments generated by routine NSLOC2.

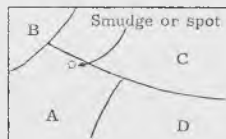
Since the frame border is always the first colour of a frame processed, it is the first entry in the output IDS. Its output face number is set to equal 1, and there is only one entry in the face table. The segment entries for the frame border are the first segments in the table, and starting at the first pointer will always point to the next segment on the right. No DDS record is put out for the frame border.

Bypass small noise faces. For each colour, all the face descriptor records are read in and stored in the input face table; all its segment records are read in and stored in the input segment table. Vertex records are read in similarly. The first task carried out in processing normal colours is to determine whether the colour can be bypassed because it is a small noise face. This may be done by checking the segments directly or by checking table COLTAB, which may have the colour as an entry because of previous calculations. If the colour can be bypassed, a branch is made to read in the next colour.

The scanner is a very sensitive device, and will record all lines on the map. However, it will also record any smudges or small specks of dirt on the map. Specks of dirt can give rise to a small face that will appear as an "interior face", as shown in Fig. A.28. ("Interior" and "exterior" are with respect to the face being processed.)

These small interior faces, if not detected and bypassed, would appear as unclassified faces in the error list. They are detected by a process (see "Area-centroid calculations" below) that checks the measure (area) of a boundary before selecting a starting segment for that boundary. If the

total measure of a boundary is  $-11 \text{ MEAS } 0$ , the boundary is a small interior face and may be bypassed.



Typical map as scanned



Enlarged view of input to PHASE4 for smudge

Fig. A.28. Small noise face (interior).

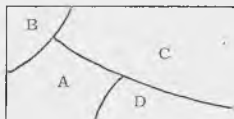
A table named BPROC is used to bypass this small interior face. The table contains entries referring to the location in the input segment table of each of the segments making up this face. In the diagram, the face appears as an interior boundary to colour A, and is composed of four segments. To bypass the boundary insofar as colour A is concerned, the switch indicating that a segment has been processed is turned on for each of these four segments. The "1" minor colours of each of the four segments making up the boundary (E, F, G, H) are then entered in table COLTAB, which contains a list of colours that can be bypassed. If two or more of the segments have the same minor colour, only one entry is made in COLTAB.

When a colour is processed, the first step carried out is to check whether it has been entered in the table COLTAB, and if so, it is bypassed. In this way, E, F, G, and H are bypassed. Before a colour in COLTAB is bypassed, however, all other minor colours of segments making up the colour are checked to see if they are entries in COLTAB, and if not, they are entered in COLTAB. Thus, when colour E is processed, it will be bypassed, but not before colour J is entered in COLTAB.

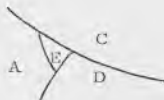
The sensitivity of the scanner and the technique used in phase 1 for colour elimination often give rise to small "exterior" noise faces at the junction of two or more lines, as shown in Fig. A.29. ("Interior" and "exterior" are with respect to the face being processed.)

This situation will also occur if the lines have been scribed on the map in such a way that a junction is thicker than normal. Phase 1 will create the

small noise face E, at the junction. This face is not classified, and will cause an UNCLASSIFIED FACE error condition that will be detected by phase 4A if the face is not deleted.



Typical map as scanned



Exterior noise face E

Fig. A. 29. Small noise face (exterior).

This condition, unlike bypassable faces, is not detectable until after the segments making up the face have been processed into the output IDS table.

After a colour has been processed, during routine COLEND, a call is made to procedure TAGUNCL in which all small unclassified faces are scrutinized to see if they can be deleted. The presence of a type "0/1" segment suggests that a large part of the face lies in the adjoining frame. Since phase 4A has no frame-to-frame communication, this cannot be checked, so it is assumed that most of the face lies in the adjoining frame and it is not set up for deletion. The minor colours of all the segments (except those concatenated in the output) making up the face are compared. If one can be found that is unique, but is not equal to 0 or 1 and is smaller than the colour being processed, then the face is set up for deletion. The face record in the IDS has its face number set to 999999 and the SDISS switch for the segment with the unique colour is set to "9". The IDS face number for the face corresponding to the unique left colour is determined, so that a supplementary (Record Code = "4002") DDS record may be created. In effect, an extra DDS record is created which causes the area of the face being deleted to be added to the area of the unique left colour (A).

The face has now been set up for deletion, but the actual deletion does not take place until the end of the frame, by procedure FRMEND. To delete the face, the tagged face and corresponding segment are located. The segment is eliminated from the IDS segment table, and pointers are adjusted to include the remaining segments of the deleted face as an integral part of the

face whose colour is the unique left colour.

Because of the complexity of the IDS, and the fact that many such faces may have been set up for deletion, other checks must be made to the IDS after deletion.

Trace boundary, concatenate segments. If the face is a valid one, a starting segment must be found. A hierarchy of search routines is employed to achieve this:

- SEARCH-A A simple search that locates most starting segments.
- SEARCH-B A complex search, invoked when SEARCH-A fails.  
SEARCH-B ensures that any segment it selects is part of a closed boundary. It is used to invoke gap-closing routines in colour 1 situations (see above), and to detect and record bypassable noise faces (see above). SEARCH-B always records exterior boundaries before interior boundaries (see below).
- SEARCH-C Finds starting segments only for unclosed boundaries, lines, and pointers.
- SEARCH-D Finds colour 0 only.

Once the starting segment for a face boundary has been located, it is processed into the output, creating IDS face and segment records as required, and moving the compact notation for the segment into the output compact notation string.

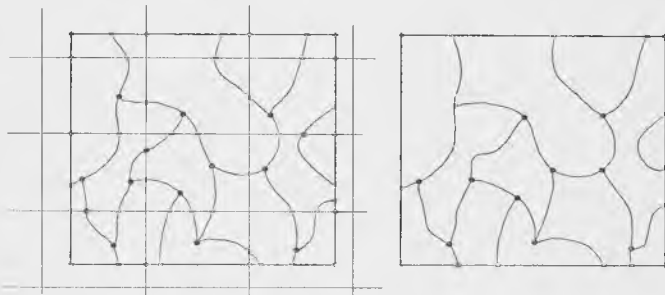
Concatenation is the process of joining two successive segments together to produce one segment. The intermediate common vertex is eliminated, and one compact notation string is created for the resulting segment.

The presence of section borders causes many of the segments in phase 3 to be concatenated because of a vertex created at the section border. The output of phase 4A (IDS) contains no vertices other than true vertices. All vertices at section borders will have disappeared.

The procedure `CONTEST` (concatenate test) is called before any segment is processed into the IDS. (This procedure is called at other times in the program, but merely to eliminate the selection of a concatenatable segment

in various situations where such a segment is not a desirable choice.) The main test for concatenatable segments is that the vertex count for the common vertex must be equal to two, and the minor colour must be the same for both segments.

The diagram (Fig. A. 30) depicts the disappearance of vertices at section borders for a typical frame.



Phase 4A input (68 segments)  
showing the 68 input segments  
with the superimposed section  
and frame grids.

Phase 4A output (37 segments)  
showing the 37 output segments  
after concatenation. Note the  
disappearance of all section  
border vertices.

Fig. A. 30. Elimination of vertices at section borders.

A colour may also consist of more than one boundary, which may be a mixture of exterior and interior face boundaries, as shown in Fig. A. 31.

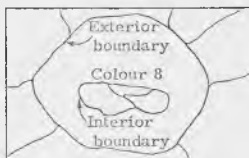


Fig. A. 31. Exterior and interior face boundaries of colour 8.



Each boundary must be completed before the next one is begun, and all the pointers and counters associated with a boundary must be initialized at the start of a boundary. Because some segment coordinates are altered and others are generated, the test to ensure that a boundary closes depends considerably on the frame list work file (table FRSEC) which cross-references the input and output tables.

Error check. When all the boundaries for a colour are determined and all the segments have been processed, the descriptor data for that face are checked before they are used to create the DDS. There should be one and only one face record for each colour that comes into phase 4A, except for colours 0, 1, and colours that are bypassed or will be deleted by phase 4A, as noted above. Several error situations may occur, in which the data for the colour are written out on the ERRORS file for inclusion in the ERROR LIST. These errors are checked for during routine COLEND, except for the last in the list below. This type of error is detected and stored when the face record is read.

<u>Error message</u>	<u>Cause</u>
MAP EXTERIOR CLASSIFIED	- Colour 1 has a face record.
UNCLASSIFIED FACE	- No face record for the colour
MULTICLASSIFIED FACE	- More than one face record for the colour.
SMALL AREA CLASSIFIED FACE	- A noise face has a face record.
FACE RECORD CONTAINS BLANK CLASSIFICATION DATA	- The classification data for a record contains all blanks.

Completion of area centroid calculations. The input segment table contains the values for the MEASURE (area) and MMX/MMY (X and Y centroid moments) for each segment. If these are totalled for all the segments making up a face, the totals yield the area and the coordinates of the centroid for the face. The area is in square grid units.

The total MEASURE for all segments making up the frame border is checked against the theoretical total for a given set of S, F, and E factors. If the two do not agree, an error situation exists and the data for the frame border (COLOUR = 0) is put out on file ERRORS, along with the message "AREA MISMATCH ON FRAME BORDER".

The totals for MS/MMX/MMY for all faces in a frame are calculated and compared with the theoretical frame totals. If the two do not agree, a message is printed on SYSOUT giving the actual and theoretical values. An error severity code of "B" is put in the map control block, but no message or face data appear in the ERROR LIST.

Build IDS records. At the end of the frame, procedure FRMEND is called. The procedure modifies the new IDS to eliminate any small exterior faces that have been tagged, corrects other small error situations in the IDS, adjusts the vertex counts, and writes out the four IDS records for the frame. The next frame is then processed in the same manner as the first.

Update MCB and write DDS. When no more frames remain to be processed, the MCB is updated and rewritten and procedure MPSGOUT is called.

The output from phase 4A is thus:

- (1) The new image data set (IDS), which is created in phase 4A and contains the image data for the frames processed in one run through phase 4A.
- (2) The new descriptor data set (DDS), which is created in phase 4A and contains the descriptor data for every face processed in one run through phase 4A.
- (3) The error list for processing in the manual error correction subsystem.
- (4) The normal run listing, containing a limited trace and any error messages pertaining to the actual running of the phase 4A program.

The following programming restrictions apply in phase 4A for a region size of 390K. The particular set of table sizes may be rearranged.

Compact notation/input segment	maximum	250 bytes
Input plus generated segments/colour/frame	maximum	700
Total compact notation input/colour/frame	maximum	7000 bytes
Number of classification records/colour	maximum	25
Output compact string/frame	maximum	15000 bytes
Output segments/frame	maximum	1200
Output faces/frame	maximum	800
Total number of input segments/colour plus total number of messages/colour	maximum	1500
Maximum number of input and generated segments/frame		2500

The following inherent distortions occur in phase 4.

Elimination of small faces and lines. Lines that form no part of a closed figure and are less than 10 points in length are considered to be noise lines, and are eliminated by phase 4. The face shown in Fig. A. 32 would thus be considered not in error.

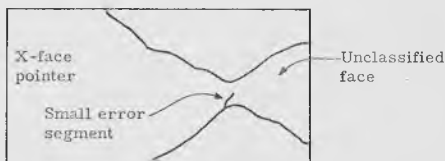


Fig. A. 32. A compound error.

The face to the left is uniquely classified and is in error only because of the scribing error shown. The face to the right is mistakenly unclassified. Because of the gap, the two faces are assumed to be one and that single face is considered not in error.

The situation depicted in Fig. A. 33 shows a further small distortion caused by dirt or widely scribed input lines. This distortion is negligible for large faces.

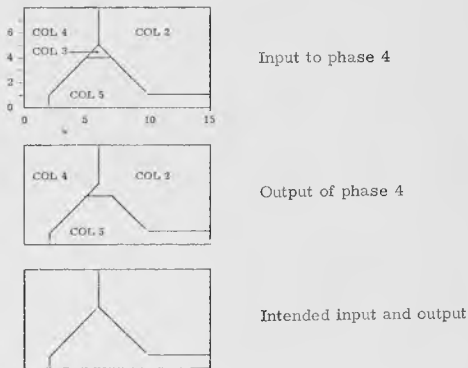


Fig. A. 33. Distortion error in phase 4.

J. White made significant contributions to the design and programming of phase 4.

#### Manual Error Correction (MEC) Subsystem

The MEC subsystem has been designed to correct errors that appear in the error listing produced by phase 4A. Certain errors are not correctable on one pass through the MEC subsystem. In particular, errors caused by program faults rather than truly illogical conditions cannot necessarily be eliminated, as the error message itself may be an error in such conditions. Similarly, errors in the phase 4A listing may be symptoms of a more complex set of errors, the components of which are only revealed when the apparent error is corrected. These conditions become less frequent as the system is developed. The description below outlines the logical operations of the MEC subsystem.

As far as MEC is concerned, phase 3 output is the source of all evil, that is, it is the source of all errors found by phase 4A and listed in the error listing. To correct the errors at their source, it would be necessary to rescan or redigitize or both, and then rerun phases 0, 1, 2, 3, and 4A. MEC is designed to create a new phase 3 output that is error free, thereby eliminating the need to correct the input to phases 0 and 1.

The sequence of logical operations that make up the MEC subsystem are illustrated in the flow chart (Fig. A. 34).

Error listings. The errors detected in phase 4A are listed and classified into categories A, B, F, G, J, and Z. Category A errors are editing and validity checks, which are not handled by MEC. Category B contains serious errors that may be handled by MEC (see comments in introduction to MEC above). Errors in categories F to J are all handled by MEC. Those in category Z are information messages for the use of MEC. A brief list of the errors that may be produced is given below.

#### R4101A - R4111A Validity errors

R4115B Map border gap. This is a serious error, usually caused by gross mistakes on the source document, though it may indicate a program difficulty. MEC can sometimes correct this error by inserting a new segment.

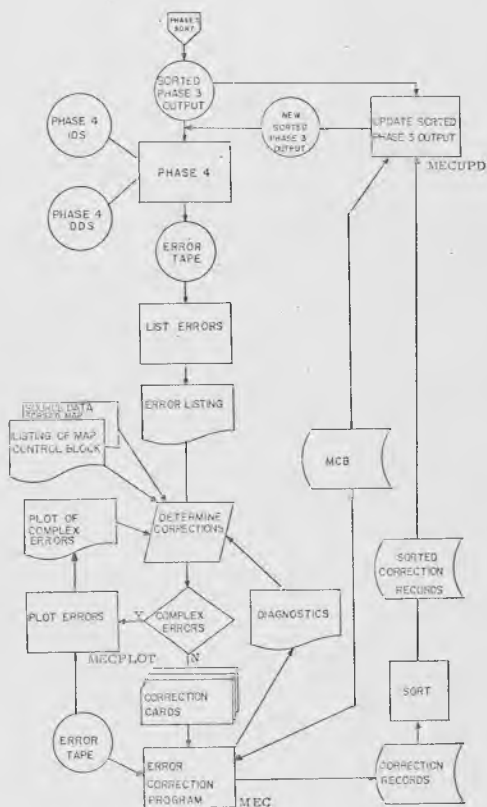


Fig. A. 34. MEC subsystem flow chart.

R4116B Area mismatch on frame border. This cannot be corrected directly by MEC, but it is usually symptomatic of other errors that can be eliminated through MEC and hence can be eliminated itself.

R4117B Unable to close exterior boundary. This implies two colours side by side with no segment separating them. Insert segment to correct.

R4118B Incorrect combination of exterior faces. A colour appearing twice on a frame. Recolour to correct.

R4160F End of line. A group of segment records with same colour on each side. Recolour to correct.

R4170G and R4173G Unclassified face. Classify to correct.

R4171G Multiclassified face. Delete or recolour to correct.

R4172G Map exterior classified. Delete or recolour to correct.

R4180J Small area classified face. Unclassified faces of less than 10 units of area are eliminated by phase 4A. Classified faces must be eliminated manually, either by deleting the face or the face record, or by changing the colour of the face record.

R4200Z - R4204Z Information messages. Help to specify above errors.

Determine corrections, optional plot. From the list of errors provided, the operator must determine and specify what corrections are necessary. This is the "manual" part of the MEC subsystem. In some complex situations, notably those requiring insertion of new segments, a plot of the segment records written on the error tape is helpful. This is accomplished by subroutine MECPLOT. The operator has at his disposal the command language of MEC, which contains the verbs: COPY, CHANGE, INSERT, DELETE, and END. These key words, coupled with definitions (and some conventions such as parentheses) are used to specify the necessary corrections:

- |        |                                                                                                                                                 |
|--------|-------------------------------------------------------------------------------------------------------------------------------------------------|
| COPY   | - A starting command which, with a subscripted definition, specifies the maps to be included in the new data set created by the update program. |
| CHANGE | - With the definition following, it specifies a change from one entity to another, for example, change of one colour to another.                |

- INSERT        - With the definition following, it specifies insertion of  
                 face records or line segments, etc., where none exist.
- DELETE        - With the definition following, it specifies deletion of  
                 face records or line segments.
- END            - End of process concerned. No subscript.

Key words that are understood by MEC and can be used in conjunction with the verbs in the grammar of description are: COVERAGE, MAP, STOP, ERROR, COLOUR, SPLIT, LEFT, RIGHT, OPEN, CLOSE, FROM, TO, POINTER, FACE, NEW, X, Y, GRID, USER, CLASSIFICATION, LINE. A full description of these terms will not be given here, but the following examples illustrate their use:

- (a) The format of a command required to add a face record is as follows:
- ```

INSERT FACE [LINE] 'line number'
          COLOUR [NEW] 'colour number'
          USER 'user face number'
          CLASSIFICATION 'descriptor data'
          [POINTER X 'X-coord.' Y 'Y-coord.']

```
- (b) The format of a command required to delete segment or face records is as follows:
- ```

DELETE [LINE] 'line number string'

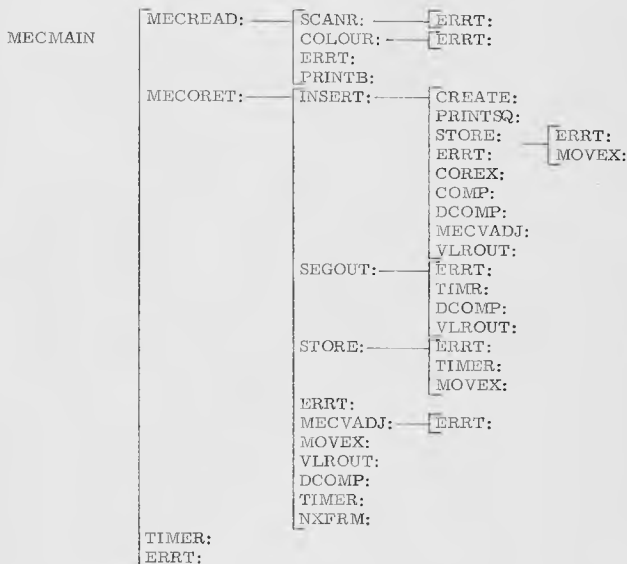
```

The operator manually determines the corrections necessary, referring to the source data, scribed map, and plot of complex errors, if produced. Cards are produced with the necessary corrections and are submitted to the error correction program.

Error correction program (MECMAIN). The error correction program contains a number of routines and subroutines.

The program is started by its main routine (MECMAIN), which allocates core, initializes the files, and calls MECREAD and MECORET. The first step is to read the correction cards prepared by the operator and store them on tables for further analysis (MECREAD). The tables are scanned (SCANR) and the key words that trigger the remaining sequence of operations are extracted. The structure of the commands prepared by the operator is

edited for validity, and in particular the directions to change or insert colours are examined (COLOUR). If errors are found in the operator's submission, then a diagnostic report is printed (ERRT) and the operator corrects his commands and resubmits them. The hierarchical calling order of the error correction routines is given below.



When the commands have passed through the initial stage of the program successfully, routine MECORET reads the error tape from phase 4A and performs the corrections "delete" and "change". At the same time, the same routine prepares the required tables and a work file for the "insert" command. To insert a segment is a somewhat more complex task than changing a colour record, as changes in the image are involved. The operation is carried out under the control of subroutine INSERT. First, the new segment required is laid out in a small block of core calculated and created for that purpose (CREATE, PRINTSQ). Then the line segments



from the error tape are conceptually superimposed on the same block of core (STORE). The compact notation manipulations are handled by sub-routine DCOMP. The new segments and vertex adjustment records are output by routines SEGOUT and MECVADJ. Routine VLROUT handles variable-length records encountered, and routine TIMER produces timings for each individual production run. If illogical conditions in the operator's commands become apparent during the INSERT, STORE, SEGOUT, or MECVADJ routines, the appropriate diagnostic messages are generated (ERRT, MOVEX) and are passed to the operator so that the correction commands can be revised. The output from MECMAIN is a set of records that are logically correct in terms of the error tape and the correction commands.

These corrected records are sorted into numerical sequence of colours within numerical sequence of frames, before being passed to the update routine (MECUPD).

The update routine (MECUPD) accepts as input the sorted output from phase 3. It compares this "master" with the "detail" produced by MECMAIN, and updates the records on a frame-by-frame basis. A new sorted phase 3 output is written and is passed to phase 4A for processing. The MCB is updated.

R. Whittaker made significant contributions to the detailed design and programming of the manual error correction subsystem.

#### Phase 4B

Phase 4B is a continuation of phase 4A and requires that the data being processed are error free, as they are to be matched with the main data bank. The phase is thus carried out after the manual error correction process has been completed.

As described above, the data being processed are now in the basic form of the data bank but the records are still based on individual maps, and numbered on a map-by-map basis. The map edges of the new data have not been matched with the map edges of the data in the main data bank. The new data have not, in fact, been merged with the data in the main data bank.

These operations are carried out in phases 5 to 8. Phase 4B sets the stage for these operations by listing the map edge data of the new records it has processed and of the abutting edges in the main data bank, in preparation for the matching operations in the succeeding phases.

The files input to phase 4B are the new IDS, the new DDS, and the existing IDS master file (IDSMAST).

The straightforward sequence of logical operations carried out in phase 4B is given in the flow chart (Fig. A. 35).



Fig. A. 35. Phase 4B flow chart.

The logical operations involved are as follows:

Initialization. This operation allocates core space to the various files, initializes the border frame record file, and performs most input/output functions.

Read new IDS and IDS master and identify border records. The frame border record includes header information, the border frame number, the adjacent frame number, the face numbers, the coordinates of start and end vertices of frame border segments (it can be assumed that a frame border

segment is a straight line), and the descriptor data for the face from the new DDS. These data are now passed to phase 5 for matching with the DDS master.

The output from phase 4B is thus the frame border record. This is sorted by map border before being passed to phase 5.

#### Phase 5

Phase 5 is designed to match the data on frame borders, primarily on map borders as noted in phase 4A above; it can also handle interior frame border data, if colour equivalences have not been completely resolved between frames in phase 4A.

Phase 5 records the information that has been sorted by map border after phase 4B, and matches faces across map borders against the DDS master file. It creates border segment records with equivalent face numbers, which are then sorted into adjacent pairs and passed on to phase 6.

The sequence of logical operations carried out in phase 5 is given in the flow chart (Fig. A. 36).



Fig. A. 36. Phase 5 flow chart.

The logical operations carried out in phase 5 are as follows:

Initialization. This routine opens the files, allocates the required areas in core, and initializes switches and counters. It also reads the MCB file.

Read input and store border record - calculate geodetics. In addition to reading the records for the main program, this routine assigns negative face numbers to all faces on the incoming maps that lie next to map borders. These are assigned sequentially, starting with -1, -2, -3, etc. This is a convenient way to identify border face numbers at this stage.

In a later phase and after equivalences have been established, each entry for a negative number will be assigned a new positive number or a pointer to another entry that has a positive face number or pointer. In this way a negative face number will be assigned a related positive face number.

During this routine the centroids are recalculated relative to the origin of GCS, and areas recorded in unit grids are converted to acres (or other specified units).

Determine equivalence. This operation matches segment records across map borders by comparing the start and end vertices of the border segments as stored in the border record file from phase 4B. The matching process can be illustrated as follows (Fig. A. 37).

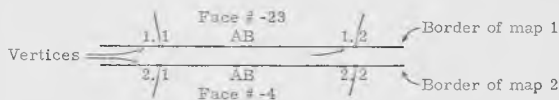


Fig. A. 37. Concept of matching.

In the diagram (Fig. A. 37) the vertices 1.1 and 2.1 match across the map borders (that is, they have the same value in the vertex table). Similarly, vertices 1.2 and 2.2 match. Given this match, the descriptors for the two faces are examined in the border data file and the DDS master respectively. In this case, the descriptors (AB and AB) match. Face number -23 can now be directly equated with face number -4 in an equivalence table.

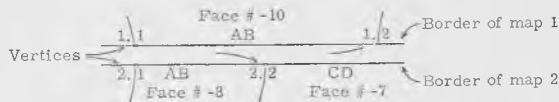


Fig. A. 38. Non-matching vertices and descriptors.

In the diagram (Fig. A.38), the vertices 1.1 and 2.1 match across the map borders. Vertices 1.2 and 2.2 do not match (their values in the vertex table are different). The descriptors match for the first part of the distance from 1.1 to 1.2, but not for the second part. Accordingly, the situation is interpreted as representing a frame-aligned segment, as shown in Fig. A.39.

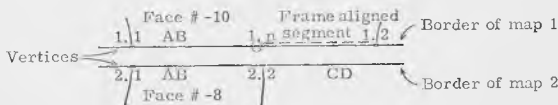


Fig. A.39. Frame-aligned segment.

A new vertex (1. n) is created and entered in the vertex table. The original segment (1.1, 1.2) has now been split into the segments (1.1, 1. n) and (1. n, 1.2). The segments (1.1, 1. n) and (2.1, 2.2) now match (their vertices match in the vertex table) and face number -10 can be equated with face number -8 along that part of the border in the equivalence table. Segment (1. n, 1.2) is designated as a "frame-aligned segment" with different face numbers on either side. If truly adjacent frames are being compared, the above process resolves the match across frame borders. Multiple mismatch indicates error in map or frame numbering.

Write face and border data. Pairs of equivalent face numbers are recorded on the border data file. This file is sorted into face number sequence before being passed to phase 6. The programming limitation to be noted is that no more than 50 maps can be processed through phases 5 and 6 at one time.

B. Sparks and P. Bedard made significant contributions to the detailed design and programming of phase 5.

#### Phase 6

The primary purpose of phase 6 is to create the records that will be used to update the DDS master in phase 7. This involves assigning resolved face descriptors to the matched records at map edges, combining them with the valid descriptor records from the new map interiors, and passing them forward. Up to this point, matching across borders has been concerned with individual segments that make up the border. Now attention is focussed on the faces. This phase accumulates the matched border segments

and assembles them by face. It recognizes that faces may be complex in shape, may wander over the map border in more than one place, and may be made up of more than one colour. This phase allows the face colour equivalences to be understood and generates instructions and data for correct record creation in phases 7 and 8. It does not actually make the changes in phase 8.

In addition, phase 6 rennumbers the faces as necessitated by new information from the incoming maps.

Information about the resolved equivalent face and border data is passed on to phase 8, so that the new IDS file can be sensibly merged with the IDS master in that phase, and map-to-map continuity ensured.

Phase 6 also updates the master coverage control block.

Phase 6 is the final stage of processing before the actual updating of the IDS master and DDS master with the new maps.

The data input to phase 6 comprises:

Header records, new DDS data, border data (map border face number equivalence and frame border face number equivalence, the latter type only being passed on when phase 4 cannot ascertain the equivalence of face numbers across a frame border within a map; phase 5 produces these records when it ascertains that the faces across the frame border are not equal), DDS master, and the master coverage control block (MCCB).

The data output from phase 6 comprises:

- (1) The DDS update to phase 7.  
This contains instructions to delete existing records in the DDS master, and information to add to the DDS master prepared by phase 6.
- (2) The equivalence file to phase 8.  
This contains: the changes (increment and decrement) needed to make the face numbers in the IDS master consistent with the DDS master, after update; two records for every border segment processed in phases 5 and 6 (and two records for interior frame border segments if between-frame equivalences are still

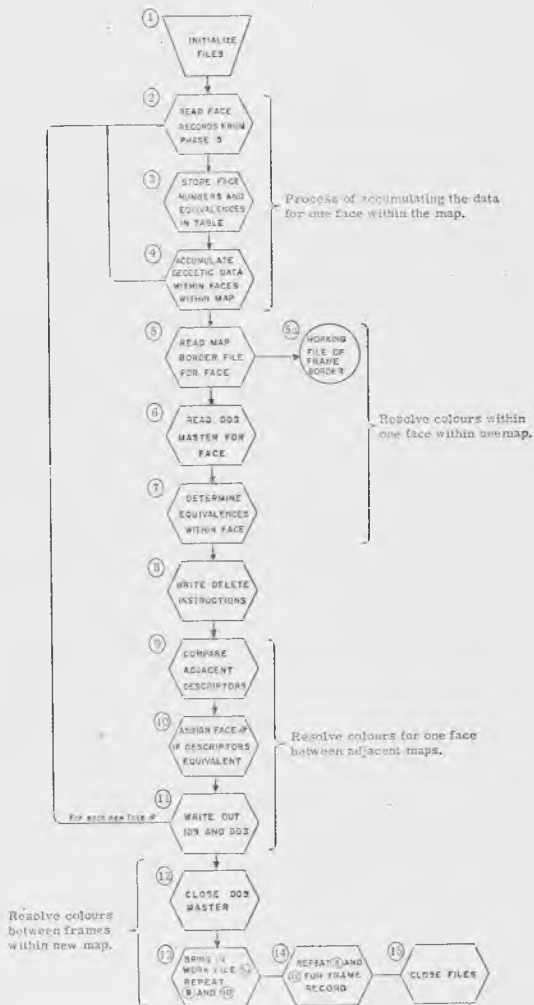


Fig. A. 40. Phase 6 flow chart.

unresolved); and changes needed to correct any face numbers for areas that match across map borders in the IDS to resolved face numbers.

The sequence of logical operations carried out is shown in the flow chart (Fig. A.40).

The logical operations involved are as follows:

Initialization. This operation allocates core space to the various files, initializes tables, and performs most input/output functions.

Read face records from phase 5.  
Store face numbers and equivalences in table.  
Accumulate geodetic data.

These three operations together are the first steps in accumulating the data from one face within one of the new maps. The geodetic data for different parts of one face in different frames are accumulated and summed.

Read map border file for face.  
Working file of frame borders.

The map border segment file is now read, and any segments that are unresolved across frame borders within a new map are placed in a separate working file.

Read DDS master for face. This completes the steps for accumulating the data for one face. All available data are now on hand to allow two types of equivalence to be observed.

Determine equivalence within face. This step recognizes that one face may have separate parts with different colours within one map.

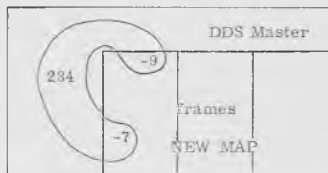


Fig. A.41. Equivalences within face.

In the diagram (Fig. A.41), the temporary colours -7 and -9 are equivalent, as they are part of the face already assigned colour 234 in the DDS master.



Write delete instructions for DDS master. Instructions are generated that will delete one of the colours on the new map and recolour the record with the second colour, to set the stage for recolouring with a new number.

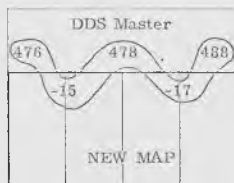


Fig. A.42. Deletions in DDS master file.

For example, in Fig. A.42, 476 and 478 would be deleted, leaving 488.

Compare adjacent descriptors.

Assign face numbers if equivalent.

Now the descriptors in the face are compared and, if they are equal, the face numbers are noted as being equivalent. The case in Fig. A.41 above is easy to resolve (number 234 would be assigned), but the case illustrated in Fig. A.42 is more complex as there are several "final" numbers already assigned in the DDS master. This may take two passes of the match operation (in phases 5 and 6) to resolve.

Write out IDS and DDS face number change instructions. Equivalence between face numbers with same descriptors is given here. The process is repeated for the next face, until all border faces on the new map have been resolved.

Close DDS master file.

Read work file of frame borders.

Write out frame border delete and change instructions. The interior faces that are unresolved across frame borders are now resolved and the appropriate delete and change instructions are written for passing on to phases 7 and 8.

The diagram (Fig. A.43) illustrates the concept of first matching over map boundaries and final matching over interior frame boundaries, when the external information has been used to resolve the former equivalence.

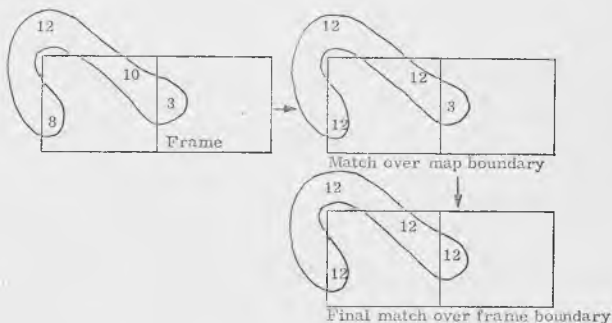


Fig. A. 43. Matching over frame and map boundaries.

B. Sparks made significant contributions to the design and programming of phase 6.

#### Phase 7

The purpose of phase 7 is to update the master descriptor data with the descriptive information related to the new maps entering the data bank. New face records are added to the DDS master, and any faces to be updated have their areas and centroids recalculated and their frame list readjusted to include all the frames in which they appear. The MCCB is also updated to reflect the new status of the DDS master.

Records input to phase 7 are:

- (1) DDS update file from phase 6, containing commands to delete records in the DDS master and information to add to the DDS master.
- (2) DDS master.
- (3) MCCB.

Records output from phase 7 are the revised DDS master file, error messages, and updated MCCB file.

The sequence of logical operations carried out in phase 7 is given in the following flow chart (Fig. A. 44).

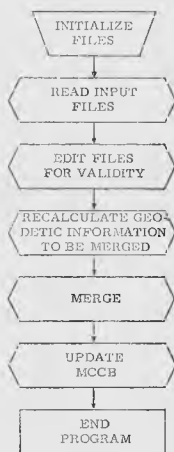


Fig. A.44. Phase 7 flow chart.

The logical operations involved in phase 7 are as follows:

Initialization. Switches are initialized and normal end-of-file procedures are set. Files are opened and tables are initialized.

Read input files for each face. The input files are read into assigned tables in core.

Edit files for validity on a face group basis. Record sequences, record types, the number of records, and the content of records are checked. Internal consistency in all aspects of the files is required before merging is undertaken. Error conditions are printed out if they have not been met.

Recalculate geodetic information to be merged. Centroids of faces at map borders that are equivalent are merged to produce a centroid reflecting the true centroid of the new face. Similarly, the areas are summed to reflect the true area of the faces to be merged.

Merge. Routine BSMERGE merges the DDS update file frame list with the DDS master file frame list on a face-by-face basis.

Write revised DDS master. The new master file is written out.

Update MCCB. The MCCB file for the data type being processed is read and is updated with the new centroid, and area calculations are added to it from the revised DDS master. Any errors in this phase can only be caused by a program or sorting error, and when they occur processing is terminated until they are corrected.

The programming limitation in phase 7 is that a maximum of 8170 frames are permitted in the new IDS.

B. Sparks and P. Bedard made significant contributions to the design and programming of phase 7.

#### Phase 8

The purpose of phase 8 is to create a revised image data set with information on the new maps entering the data bank. Map borders are eliminated as maps are joined. Face numbers of faces crossing the borders are resolved. Any necessary segment splitting is done in a similar manner to segment splitting in phase 6. The vertex counts along the border are set and the MCCB is updated to take into account the revised status of the IDS.

The inputs to phase 8 are as follows:

- (1) The equivalence file records from phase 6, which contain:  
Change records, equivalence records, and the negative face number record (the latter to change all negative face numbers that have not been processed in phase 8 so that they may reflect the true face number).
- (2) The new IDS records from phase 4, which contain:  
Header, face records, segment records, and compact notation records for the new data.
- (3) The IDS master file.
- (4) The MCCB.

The outputs from phase 8 are the revised IDS master file and an updated MCCB.

The sequence of logical operations in phase 3 is given in the flow chart (Fig. A. 45).

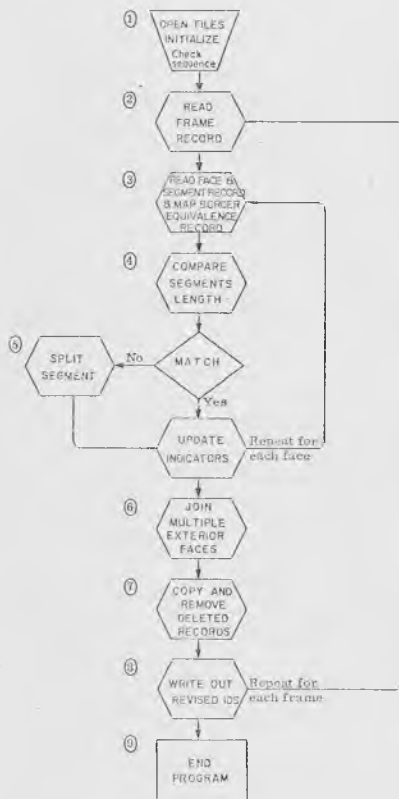


Fig. A. 45. Phase 3 flow chart.

The logical operations in phase 8 are as follows:

Initialization, check sequence, and edit for validity. This operation allocates core space to the various files, initializes tables, sets switches, performs most input/output functions, and edits input file contents for format, validity within set limits, and internal consistency.

Read frame record.

Read face and segment record and map border equivalence record.

Operating on one frame at a time and on one face at a time, in sequence, each segment is examined. Border segments are identified in the IDS and are compared with the resolved border segment descriptors being passed from phase 6.

Compare segment length. This is conceptually the same process that was applied to border segments described in the DDS in phase 6, but now the segments are taken from the IDS, that is, the equivalences resolved in the DDS are now passed to the IDS. As in phase 6, the vertices are examined. If there is a match, the indicators contained in the segment records are switched to indicate a matching segment. If the vertices do not match, the segment is split.

Split segment. As in phase 6, segments are always split, rather than extended, to create a vertex match. Given two segment records that match only along part of their length, the common part is identified, the larger segment is shortened by changing one or more of its vertex records to equal those of the short segment, and a new segment record is written to describe the piece that is split off. The new segment record will be put into the table with the others but will be labelled for one side only; this will be altered to the correct value when the adjacent frame is processed. The process of comparing segment lengths and splitting segments to achieve matches is carried on until all segments around the frame are identified with the face numbers on either side of them, or are identified as frame-aligned segments and are labelled as such so that their contribution to faces in adjacent frames can be understood.

Join multiple exterior faces. This is the process of determining the equivalences with a face that may have separate parts with different colours

within one map. Using the equivalences on the equivalence file records passed on from phase 6, the separate parts are identified as belonging to the same face. The concept is illustrated in Fig. A.46.

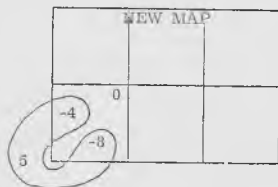


Fig. A.46. Joining multiple exterior faces.

The new IDS records for the above map would show that face number -4 was made up of segments 1, 3, and 8 (for example) and the face number -8 was made up of segments 9, 12, and 14 (for example).

The equivalence file records would show that face -4 = face 5 and that face -8 = face 5. The two faces (-4 and -8) would then be joined by creating a record that showed that face number 5 in frame 0 is made up of segments 1, 3, 8, 9, 12, and 14. In effect, the two negative numbers have been replaced by the higher positive numbers, and the segment (and compact notation) records have been concatenated behind the new single number.

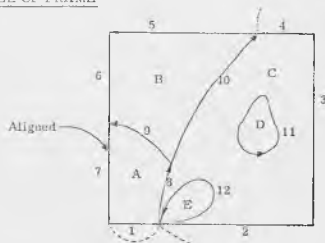
Copy and remove deleted records.

Write out revised IDS.

These two operations revise the records in light of the changes brought about by recognizing equivalences. The first step simply traces through the files and removes the redundant records. The second step writes out the complete IDS in final format using the new information. The process is repeated frame by frame in the IDS. As the data have been made equivalent between all frame borders, new and old, then the map borders have been effectively eliminated and exist only as frame borders in the final data bank. The format of the IDS for a sample frame is given in Fig. A.47.

End program. Errors in phase 8 can only be program or sorting ones. When they occur, processing is terminated until they are corrected.

## IMAGE OF FRAME



## FACE RECORD

	#	#1st Segment	Switches		Touches frame
			Multiple exterior	Interior	
1.	A	1L	No	No	Yes
2.	B	9R	No	No	Yes
3.	C	2L	No	Yes	Yes
4.	D	11L	No	No	No
5.	E	12L	No	No	Yes

These numbers are not stored in computer; records are just kept in sequence.

Switch indicates way lines traced in phase 1.

## SEGMENT RECORD

All bit switches											
Line	Next L	Next U	Next V	Next W	Next X	Next Y	Next Z	Next A	Next B	Next C	Next D
1	2R	3L	4L	5L	6L	7L	8L	9L	10L	11L	12L
2	3R	4L	5L	6L	7L	8L	9L	10L	11L	12L	
3	4R	5L	6L	7L	8L	9L	10L	11L	12L		
4	5R	6L	7L	8L	9L	10L	11L	12L			
5	6R	7L	8L	9L	10L	11L	12L				
6	7R	8L	9L	10L	11L	12L					
7	8R	9L	10L	11L	12L						
8	9R	10L	11L	12L							
9	10R	11L	12L								
10	11R	12L									
11	12R										
12											

The numbers are not stored in the computer. The records are just kept in sequence.

\*means go back up to face record and get the next face (B, C, D, etc.)

Number of segments entering into vertex minus 1 so that the maximum (7) can be represented in three binary bits.

Sample lines filled in. All lines filled in actual record.

## COMPACT RECORD

1011010110111 etc.

1 2 3

Fig. A. 47. Sample of IDS format.



The following modifiable restrictions apply to phase 8, as currently implemented:

Number of faces per frame	200
Number of segments per frame	1000
Number of compact bytes per frame	10000

B. Sparks and P. Bedard made significant contributions to the design and programming of phase 8. A. Benjamin, P. Kingston, G. Morton, and F. Jankaluk made significant contributions to the detailed design and programming of the data base format.

#### DATA RETRIEVAL SUBSYSTEM

One of the earliest design criteria for the CGIS was that, to the greatest extent possible, the manipulation and retrieval of data from the data bank should be initiated by commands that could be understood by a wide group of potential users and that could be used without extensive (and expensive) computer programming. That objective has been gained to the extent that a person with a basic knowledge of one high-level computer language (PL/I) can assemble English-type key words (from the command language) into a simple program that allows even extremely complex operations to be carried out without further human interaction. Most recent graduate students of geography would have no difficulty in preparing a request. Also, any desired manipulation or retrieval can be built into a loadable program and assigned a key word, which can then be used in subsequent programs (that is, added to the command language). However, this step may be more easily carried out by a programmer. The whole process operates under the control of a "monitor" system that interprets the key words, assembles the program that the computer actually uses, submits the result to the computer for processing, and receives the results. P. Kingston was responsible for the overall design of the CGIS data retrieval subsystem.

The relationship between a nonresident monitor system, a command language, and the resident operating system of the computer under which all operations are carried out is as follows: when a monitor system is used, it is the executive director of all normal operations within the computer, taking over manual control tasks. A command language can

operate without a monitor system, but if there is no command language the monitor system has to function as an operating system. The latter is essentially a computer-oriented control system; it guides work through the computer in a way that makes best possible use of the machine's resources, rather than in the most convenient way for the user. An operating system can be thought of as a subset of a full monitor system.

As a guide to understanding the actions of the monitor program, it is useful to consider the process of manually assembling a normal program. Usually a "source deck" of instructions is prepared by a programmer, in either machine-oriented or people-oriented code, that prescribes how the data to be submitted must be processed. The source deck is placed next to a deck of "job control language" cards that contain the set rules for handling the source deck, and the two are processed ("compiled") by the computer to provide a new set of cards ("object deck") in machine language. The object deck is still independent of input and output control routines, other pre-written subroutines (such as sine and cosine functions), and the location in the computer memory in which it will operate. The object deck is linked to the appropriate prewritten subroutines, which are already in object deck form, in a manner prescribed by control cards under the direction of a "linkage editor" program. The result is a "loadable program", but one which has not yet been assigned a specific memory location for processing by the computer. A final set of job control language rules is placed after the loadable program; it identifies the data files to be accessed by the program during processing, and instructs the computer's operating system, under the control of a program named "LOADER", to assign specific memory locations to the program.

The CGIS retrieval facilities allow each of these levels of program compilation to be intermingled in the process of an inquiry. In his initial statement, the user can thus command a combination of source decks, object decks, or loadable programs. The source decks are special-purpose requests, unique to the user concerned and written by him in either PL/I or assembler language. The object decks or loadable programs are routines already in the libraries of the system, and are called up by key word in the user's

program. The job control language is similarly stored in the system library, and provides the rules and structure within which the machine-readable requests are generated.

A general schematic of the CGIS retrieval monitor is given in Fig. A. 48.

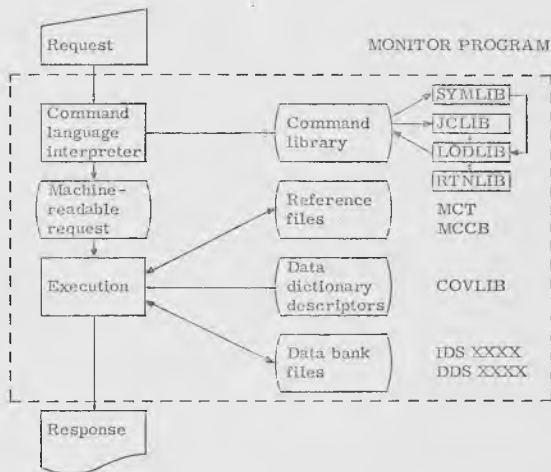


Fig. A. 48. CGIS retrieval monitor system.

The entire retrieval system is written in the PL/1 programming language and is supported by assembler language subroutines. The user codes a series of commands in the custom PL/1 of the command language and submits them to the monitor program (see diagram above). The command language interpreter translates these commands by referring to the command library, and produces a machine-readable request.

The request, still in the job control language generated by the monitor, next has to be interpreted and revised in terms of the data on which it has to operate. The reference file contains two data sets that are referred to at this stage. The master control table (MCT) is a list of all data types

(coverages) currently resident in the CGIS data bank. It provides the relative location of the control information on each coverage in the master coverage control block (MCCB), which, it will be remembered, was generated and continually updated by the data reduction subsystem. The MCT is in fact an index to the MCCB. The MCCB contains such vital information as block size, logical record length, physical location of the files mentioned in the request, the various types of single coverage in any one combined coverage, F factor, S factor, E factor, and other information such as the number of descriptors used in the coverage. The monitor selects the appropriate data needed and calculates the instructions that must be added to the request. Simultaneously, the coverage library (COVLIB) is consulted. COVLIB contains the descriptor data fields for each coverage as a set of PL/1 declarations. When needed, these are incorporated in the request.

The request is now executed, and operates upon the data bank files (DDSXXXX) or (IDSXXXX), or both, as necessary. (The XXXX in the name simply indicates the code number of the particular data set being used.) In practice, the monitor determines from the list within itself which files are required and then provides the operating system with the appropriate control information to acquire the files. The operating system determines if the needed files are loaded on the computer already, and if not it will print a message on the operator's console instructing him to get the files from storage and load them on the appropriate reading device. The entire retrieval process utilizes the resident OS/360 operating system control, and uses operating system facilities normally available in IBM 360/65 computers. While the program is being executed, the monitor returns to "wait" status (to handle requests for other tasks).

At the completion of the run, the program sends a message to the monitor indicating that the task has been completed. The monitor reads the identifier of the task and removes it from its table of active tasks. Based on that record, it does one of several things: (1) passes the output to an output device; (2) terminates the operation; (3) repeats the whole cycle. The most significant characteristic of a monitor system is that it allows one computer program to control another, based on data provided by the second one.

P. Kingston, B. Ferrier, and M. Doyle made significant contributions to

the design and programming of the monitor system.

#### COMMAND LANGUAGE INTERPRETER

The key to the monitor system is the command language interpreter, and its operation will be described briefly before the CGIS command languages are examined.

The command library at the disposal of the CGIS command language interpreter is made up of four data sets:

- (1) Source program library (SYMLIB). This library contains a set of PL/1 source programs for each "assessment language" subcommand.
- (2) Routine program library (RTNLIB). This library contains all the pre-assembled subroutines stored in object level form.
- (3) Loadable library (LODLIB). This library contains all the custom-written programs in loadable form ready for execution. These programs are called by commands in the "manipulation language".

It will be noted that the three data sets above represent three levels of program compilation: source, object, and loadable.

- (4) Job control language library (JCLIB). The job control language rules that are required to execute each command are stored in this library as a sequence of one or more job steps.

When the request is passed to the monitor for processing, the command language interpreter first reads the request. On the basis of the key word code, it searches through the appropriate file in the command library for an equal condition on an associated set of rules in the JCLIB. It calculates the interactions between various data and information requirements and the request, that is, it creates the set of suboperations to be performed to carry out the command, then it takes the result and moves it to the next stage.

If the original request was extremely simple, this grossly oversimplified description of the operation might be sufficient. A more usual, typically complex, request requires a much more complex response from the command language interpreter. A slightly more complex request might be handled as follows. First, the command language interpreter reads the

request, reads the rules in the JCLIB, and searches through the appropriate library (SYMLIB, RTNLIB, LODLIB), as before. Then, in order to process or translate the request, it must operate upon the request according to the set of rules; at least one operation has to take place for each rule in the set. The specific operation depends on the command being interpreted, for example, a simple logical comparison to determine whether the parameters are all present before going on to the next operation, or it may be a logical calculation to determine the relative storage address for the instruction being generated.

While these operations are being performed, the parameter data references and instruction labels that are encountered, their definition, and any addresses that have already been assigned to them are copied into a storage table (R). The partially translated instructions are stored in another table (J). When R has been sorted, addresses are assigned to the labels and parameters. The table R is conceptually compared with table J and the parameters in the latter are replaced with addresses from table R. As each request statement is processed in this way, it is moved to a storage medium as a machine-readable request for further use. Within the CGIS, commands that require straightforward processing are termed "normal" commands and those that generate more complex interpretation procedures are called "generator" commands. The difference is transparent to the user.

The CGIS actually recognizes two interrelated command languages. In one, the literal commands are prefixed by a # sign and belong to the "manipulation language". These are high-level commands (key words to quite complex operations), and they are interpreted by access to the JCLIB and the LODLIB.

The first command in the manipulation language is #ASSESS. This allows the combination and use of commands prefixed by a \$ sign, which belong to the "assessment language". The assessment language commands are minor versions of the overall command language structure, and call up a wide selection of specialized subroutines. The #ASSESS command can be used in conjunction with these minor commands to describe a procedure specific to the user. Thus, not only does the user have the option of employing main

key words to specify his needs, but he can, to a considerable extent, write his own special programs and load them under the #ASSESS key word. (In addition to this, the user can, of course, step outside the command language entirely, program his requests in plain PL/1 or assembler language, give them a key word, place them in a command library, and call them by key word thereafter.) Assessment language commands are interpreted by access to the SYMLIB. Both manipulation language and assessment language commands, after interpretation, may use subroutines from RTNLB.

Examples of requests. Two sample requests are given below.

(1) Information Required

How much land greater than 10 acres in area within Rideau County is now being used to grow potatoes? List the areas with their township and acreage.

Request

# SELECT	PROV,I,NEWNAME=RIDO	1
	COUNTY=RID	2
# COMBINE	RPLU=PRLU+RIDO	3
# OVERLAY	RPLU	4
# ASSESS		5
	\$ CFILE,RPLU,I	6
	ON ENDFILE(DDIRPLU) GO TO END:	7
	TTLAREA=0;	8
READ:		9
	\$ READ,RPLU	10
	IF USE='FARM'THEN	11
	IF CROP='POTA'THEN	12
	DO;	13
	\$ACRE,RPLU	14
	IF AREA = 10 THEN	15
	DO;	16
	PUT SKIP LIST (COUNTY,TWP,USE,CROP,AREA);	17
	TTLAREA=TTLAREA+AREA;	18
	END;	19
	END;	20
	GO TO READ;	21
END:		22
	PUT SKIP LIST('TOTAL POTATOES FARMED IN	
	RIDEAU COUNTY=',TTLAREA,'ACRES');	23
# MODMCCB		24
	\$ DELETE RIDO,FACTORS,FILES	25
	\$ DELETE RPLU,FACTORS,FILES	26

Explanation of Request StatementsStatement Function

- 1 & 2     Selects a subset coverage from the political coverage PROV for the county with the coded name RID and calls it RIDO. The 'I' in statement 1 is an abbreviation for inclusive.
- 3         Creates an entry in the system library for coverage RPLU made up of present land use coverage PRLU and subset coverage RIDO just selected.
- 4         Causes an overlay operation to take place to create the coverage described by 3.
- 6         Calls in the assessment program generator which creates an assessment from statements 6 through 23.
- 6 - 23     A simple assessment program, lists the areas required and finds the total required.
- Note the use of the special instructions #CFILE (statement 6), \$READ (statement 10), and \$ACRE (statement 14) in addition to the regular PL/1 statements.
- 24         Calls in the programs to modify the system libraries,
- 25 - 26     deletes the temporary coverages created, and releases their files, which have been saved in case it is desired to run several assessments on the same data or to save them for another future use.

(2) Information Required

Find an area suitable for recreation (e. g., campsites) at least 2 square miles in size, as near as possible to a lake of at least 3 square miles and to Highway 183, that is within 50 miles of the town of Renfrew.

Request

```
# GENERATE CIRCLE, CENTRE=42382811, RADIUS=50MI      1
# COMBINE RENF=PRLU+ROAD                               2
# OVERLAY RENF, GENR, LINE                             3
# COMBINE CAMP=RENF                                   4
# ASSESS                                              5
```



### Explanation of Request Statements

#### Statement    Function

- 1      Generates a circle around the town or Renfrew so that only that area will be selected.
- 2      Sets up an entry in the system library for a new coverage 'RENF', to be formed by combining the data of present land use with road or highway coverages.
- 3      Creates the new coverage RENF as defined by 2, using a selection routine for generated coverage (the one generated in statement 1) and including line data (road maps).
- 4      Sets up an entry for a new coverage CAMP to be formed from coverage RENF. Before another command refers to the coverage, an entry must be generated for it.
- 5      We now have the data needed but we must exclude any lakes or recreational areas smaller than those required. This is done by a short assessment program that will write out a new coverage consisting of only the data with the desired qualities.

Note:      The area could have been selected manually from a list output by the assessment program.

### COMMAND LANGUAGES

The suite of commands in the manipulation language and assessment language is given below. P. Kingston, K. Ward, B. Ferrier, M. Doyle, J. Thacker, F. Jankaluk, H. Knight, and P. Hatfield made significant contributions to the design and programming of the retrieval commands.

#### Manipulation Language - #ASSESS Command

# ASSESS. This command signals the start of a PL/1 program, using the assessment language commands given below. The program is first processed by the assessment language generator, which resolves all subcommands that may be present in the input (immediately after the command), and then passes this resolved input on to the OS/360 catalogued procedures for compiling, link editing, and executing the assessment program. All required job control language and necessary subroutines are automatically generated and included by the assessment generator.

There are several options available to the user coding this command, including the PL/1 compiler options, a command to the assessment generator indicating the card image to be printed, and start/stop indicators for an overlay. The defaults for this command are any PL/1 compiler options that were set as defaults when the operating system was last initialized; however, processing of compile-time macros will always be specified.

#### Assessment language.

\$ ACRE. This subcommand calculates the area in acres of a DDS face of a specified coverage.

\$ CFILE covg (1)/(0). This macro provides a PL/1 structure declaration for a DDS file of a specified coverage. It is a convenience command for addressing coverages.

\$ CLASSIFY cvg2(FROM cvg1)c-ptr. This macro writes a descriptor record for a coverage processed by the \$DISSOLVE macro below.

(FROM)  
\$ DISSOLVE cvg2 (EXCL) cvg1 c-ptr. This subcommand tags a face to be coalesced or merged with adjacent faces that have the same output descriptor. It is the only subcommand available in the assessment language for modifying the image data set. It is used in conjunction with CLASSIFY if it is wished to change image data by removing boundary lines. FROM = change specific. EXCL = change everything except.

\$ EQUATE cvg2 FROM cvg1. This subcommand moves all DDS face record contents, except for the descriptor data, from one area to another, one face at a time.

\$ INITP (name-1, name-2). This function generates a PL/1 "PROC OPTIONS (MAIN);" statement, that is, it adds a title to a custom PL/1 program.

\$ READ covg(F). This function reads and checks a data bank DDS face record for a specific coverage.

\$ SELECT cvg2 FROM cvg1. This function creates a subset data bank IDS based on a list contained in a DDS face record or records. It copies an IDS face.

\$ SQMI covg. This function calculates the area in square miles of a DDS face of a specified coverage.

\$ TERMP. This is an optional way to terminate a PL/1 program written using #ASSESS. Its function is to generate the PL/1 "END;" statement.

\$ WRITE cvg2 (FROM cvg1). This function writes a data bank DDS record with a specified coverage code.

#### Manipulation Language

```
# CIRCLE (covg, )CENTRE=DDD, MM, SS, DD, MM, SS
          , RADIUS=9999.999
          (, POINTS=999)
          (, SFACT=99)
          (, FFACT=99)
```

This command allows a user to generate a circle in a data bank format, that is, an IDS containing the image of a circle. The user can define the circle with a given centre, a given radius, and a given number of points on the circle. The circle is defined as an inscribed polygon. This circle is then available for overlaying on top of any coverage, and thus defining a subset coverage.

# COMBINE cvgc=cvg1+cvg2+cvg3...+cvg8(, M). This command is used to satisfy the prerequisites for the OVERLAY and MERGE commands. Its function is to add entries to the master coverage control block and COVLIB for a new coverage that is to be a combination of already existing coverages. If (, M) is not specified, the concatenated data descriptors for each coverage would be the same and only the first set is used.

# DDSGENR covg plus data cards containing face number and descriptor. The purpose of this function is to generate a data bank DDS containing only face descriptor data. The program is generally used to satisfy the prerequisite of #OVERLAY that there be a matching DDS for every IDS. (This command is normally used in conjunction with #POLYGON and #CIRCLE.) It can also be used to generate a mock DDS for testing purposes.

```
# DDSLT (NOPRINT)(, COPY)(, NONEWNAME)
          (PRINT)(NOCOPY)(NEWNAME=covg)
```

This is a utility command whose purpose is simply to process the DDS. It will print selected fields, copy an entire DDS, and rename a DDS from one coverage code into another.

# MERGE covg (, T)(, V=SER=vol-id)  
(, D)(, V=REF=vol-id)

The function of this command is to create a new IDS/DDS pair in data bank format, containing the IDS and DDS of non-overlapping component coverages merged and renamed. No actual processing takes place and the final result is equivalent to an execution of the O/S sort/merge utility program. T specifies output on tape. D specifies output on disc.

#MODMCCB. This is another utility program, whose function is to manipulate the master coverage control block and the associated COVLIB entries. The command makes use of subcommands and optional data cards. All subcommands are similar to those in the command #ASSESS, in that card column one must be a \$ sign.

\$ADD covg(FACTORS)(V=unit=vol-id) -MCCB (and COVLIB) cards-	Add a coverage to the system library
\$UPDATE covg(FACTORS)(V=unit=vol-id) -MCCB (and COVLIB) cards-	Change the contents of the system library for a coverage
\$DELETE covg(FILE)(FACTORS)(ONLY)(V=unit=vol-id)	Remove a coverage from the system library and from the <u>data bank</u>
\$LIST (covg)(FACTORS) (MCCB)	List the system library entry for a specified coverage
\$COPY cvg2 FROM cvg1(V=unit=vol-id)	Create a new entry the same as an existing entry
\$SAVE	) Create back-up files.
\$RESTORE	

#POLYGON (covg)(, SFACT=99)(, FFACT=99) plus data cards containing coordinate list. This command is similar in function to #CIRCLE, but it generates an image data set in data bank format containing the image of a polygon. The polygon is defined by a coordinate list supplied by the user. The only requirement is that it must be a regular polygon, a single complete figure enclosing a single area, with no sides merging or crossing.

#SELECT (EXCLUDE1)(, NEWNAME=covg)  
(INCLUDE)

The purpose of this command is to produce an image data subset in data bank format, by including or excluding records from an input image data set on the basis of a list supplied by the user. It is a similar function to \$SELECT and conceptually the reverse of #MERGE.

#SORTDDS. The purpose of this command is to sort one or more descriptive data sets in data bank format into ascending sequence by face number, record type, and coverage code. It executes the O/S sort/merge utility program. All JCL overrides must be provided by the user.

The final command in the manipulation language is the one that calls the overlay function. It is an important command as it greatly expands the manipulative capability of a multi-data-set data bank. To demonstrate the types of logical operation brought into play when the above commands are employed, and to illustrate the overlay process itself, a brief description of the overlay logical operations is given below.

#### #OVERLAY

The #OVERLAY command enables the user to overlay one coverage directly on top of another, producing a third coverage that is the merged image of the two inputs. Up to eight input coverages can be overlaid in one operation. The image data of the input coverages are laid one on top of the other to form a combined image, the descriptor data for the related image portions are concatenated, and the geodetic data, area, and centroid are recalculated to reflect the new faces.

The #OVERLAY command causes the generation of a subsystem that is made up of job control language encompassing the two programs #IDSMERGE and #OVLAY. The related functions of these two programs are illustrated in Fig. A.49.

The #IDSMERGE program is a simple merging operation in which the IDS coverages to be merged are placed on one file in sequence, according to frame number and coverage code. This is a straightforward operation and simply presents the data to the main #OVLAY program in a convenient manner.

The overlay process itself is accomplished in two steps. The first carries out the image overlay, produces the overlaid image, and passes information on resolved face equivalents on to the second step. The overlay of the descriptor data set makes use of the selected DDS files and the information from the first step, and produces the overlaid descriptor data set.

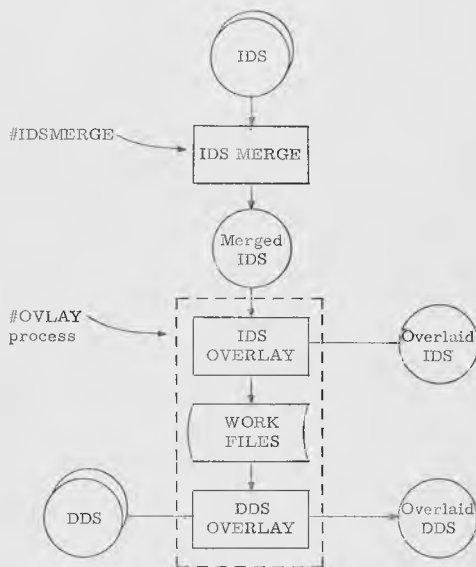


Fig. A.49. Conceptual diagram of #OVERLAY command functions.

The sequence of logical operations in the IDS overlay step is shown in the flow chart overleaf (Fig. A.50).

The logical operations involved in the IDS overlay are as follows:

OVL. The OVL program first edits the data and the request. The master coverage control block (MCCB) is checked for conditions that would make the overlay invalid, and error messages are generated for conditions that are not acceptable. The input coverage names are sorted into collating sequence and listed on the output printer for a final check by the operator. The F and S factors for each coverage involved are taken from the MCCB and checked for validity. The core requirements of the overlay program are calculated, various options being worked out to find the most efficient

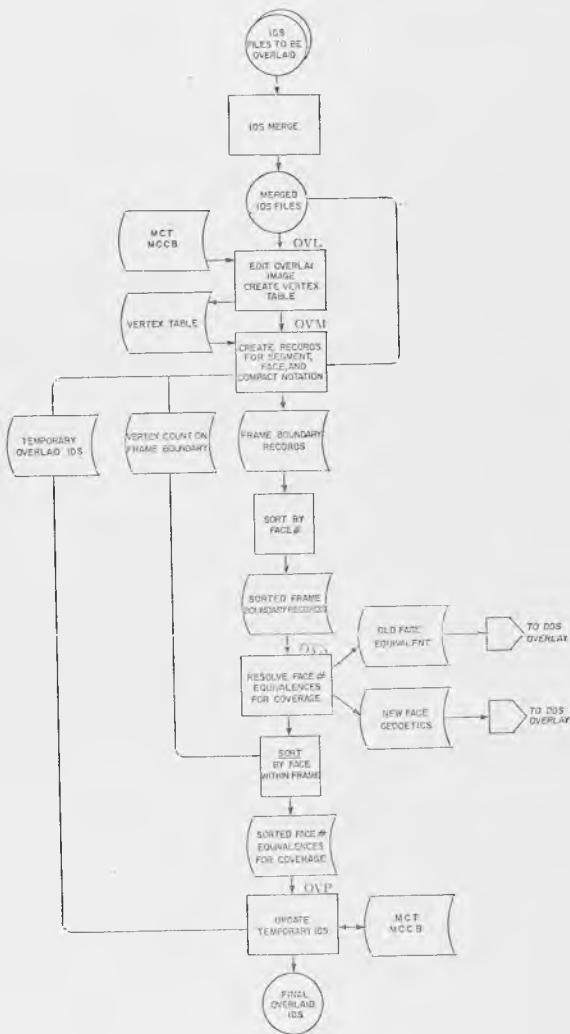


Fig. A. 50. Flow chart showing sequence of logical operations in IDS overlay.

method of using the available core space. An indicator of the calculation option to be used is set, to be referenced during processing. The core available, the core requirement, and the processing option are also printed out.

The IDS data bank coverages are read in on a coverage-within-frame sequence, as put out by the #IDSMERGE step. For each common frame group, a "bit image" of the frame is laid out in core. Essentially, a blank area of core adequate to handle every point within the frame is allocated, and all the image records from common frames are laid out in that space. The vertices for the line segments are plotted and edited for validity. The compact notation for each segment is then traced from vertex to vertex. This is carried out until all line segments from all IDS coverages being overlaid are contained within a single frame record. All points where segments intersect are now recorded in a vertex table, together with the original open and closed vertices for each input segment, points of segment merges, and representative points for diagonal intersections. This table of vertices is sorted into sequence, equal entries are eliminated, and if it has been necessary to sort the table into groups to make optimum use of core, the sorted groups are merged together. The final output, with certain calculated values and editing information, is passed to OVM.

OVM. On a frame-by-frame basis, the sequenced vertex tables output by OVL are read in, as well as the IDS for each frame of each coverage. The segments of each input face are traced and the face number is attached to the correct side of the segments. Each segment is retraced, checking for the intersections recorded in the vertex table, generating a new segment at each recorded vertex, and recording the new segment number and direction of intersection in the vertex table. This is done for each input coverage for the frame. The similarity between this and the boundary tracing carried out in phase 1 of the data reduction subsystem is evident. The whole overlay process is, in fact, a logical copy of phases 1 to 8 of the data reduction subsystem, with the assumptions that the data being handled are error free and already in the form of single-point-width line segments.



The new segments are now linked to one another by pointers, produced by tracing from vertex to vertex and outlining each area or face with unique characteristics (combinations of input faces from each coverage). A colouring process identifies all the input coverage faces that enclose or form the new faces by passing input face numbers across segments, after all the segments bounding the face have been checked.

Interior faces not touching an exterior portion are identified and marked.

Segments are concatenated wherever possible, where only two meet at a vertex. Geodetic data are calculated for the new segments.

Geodetic totals associated with an area and X-Y coordinates for each face within the frame are found and information is passed to OVN for interframe linkages. One record is put out for each face, with colouring data (input face numbers) and geodetic data. Also, records for each vertex and segment on frame borders are passed to OVP and OVN respectively.

A temporary IDS in data bank format on an independent frame-by-frame basis is passed to OVP.

OVN. The frame boundary records put out by OVM are sorted by input face number before being processed by OVN. This groups together the records for multiframe faces. The OVN program assigns a new face number to each group, totals their geodetic data, and converts them to acres. Face centroid coordinates are calculated relative to the GCS origin. The first product is a geodetic transmittal record which is, in effect, a dummy DDS record, since it has no descriptive data at this point. The new face numbers are passed to OVP after being resorted into a sequence of face numbers within frames.

Face equivalence records indicating which input face numbers form the output faces are written on a work file to be passed to the DDS overlay operation, so that the missing descriptor data can be extracted from the input DDS files and added to the geodetic data.

OVP. The sorted output from OVN is read, along with the temporary IDS output from OVM. If necessary, for the input records, segments on the temporary IDS are split into two or more parts, as indicated by the vertex

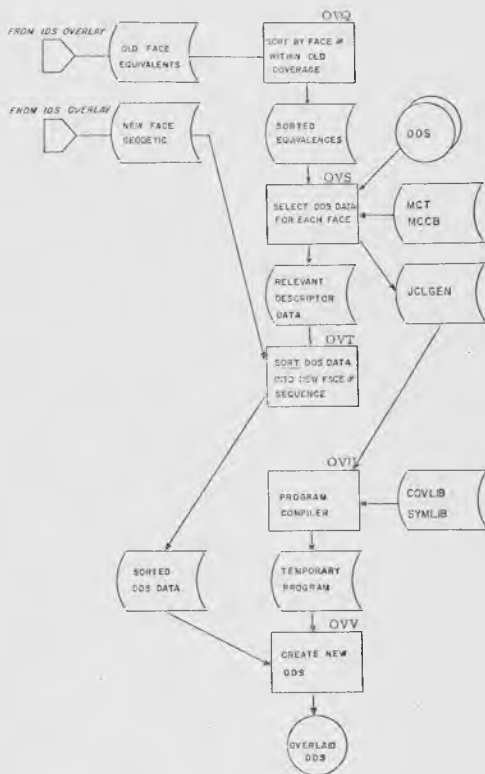


Fig. A. 51. Flow chart showing sequence of logical operations in DDS overlay.

records. (This is the familiar segment-splitting process carried out in phases 6 and 8 of the data reduction subsystem.) The vertex counts are also updated for these vertices on frame borders, to allow for segment intersections on the opposite frame borders. Also, alignment is set for segments not previously known to be aligned, and alignment is removed for segments that were aligned and should not have been. The new face numbers are assigned, multiple external faces for reentry of the face into a single frame are recognized, and face-aligned segment records are created for segments aligned with the frame borders. As with phase 8, the output of the program is a new coverage IDS. The MCT and MCCB records are updated with information describing the newly created IDS coverage.

The sequence of logical operations in the DDS overlay step is shown in the flow chart (Fig. A.51).

The logical operations involved in the DDS overlay are as follows:

OVQ. The face equivalence records arrive at OVQ as a series of face numbers for the faces created by the image overlay process, each of which has attached to it the numbers of faces from the original coverages. This is conceptually illustrated in Fig. A.52 below.

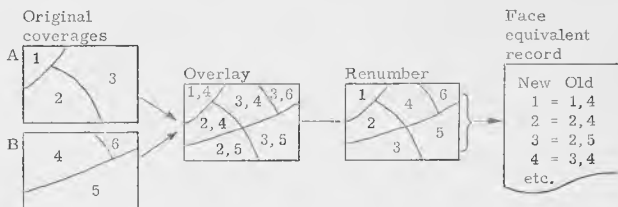


Fig. A.52. OVQ process.

OVQ simply sorts the face equivalent record into a sequence of face number within original coverage to make it easy for the next step, OVS, to select

the descriptor data from the original coverage files (Fig. A.53).

New	Old	Old	New
1	=	A1	= 1
2	=	2	= 2, 3
3	=	3	= 4, 5, 6
4	=	B4	= 1, 2, 3
etc.		5	= 3, 5
		etc.	

Face equivalence record

Sorted face equivalence record

Fig. A.53. Sorting of face equivalence record.

OVS. Using the sorted face equivalence records passed from OVQ and the input coverage DDS files, this operation extracts from the latter the face descriptor records needed to form the face descriptor records of the final overlaid DDS.

OVS also refers to the MCCB for parameters and, in conjunction with a prewritten routine in SYMLIB named DDSOVL, generates a PL/1 source program. After being compiled in OVU, this program will be used to create the overlaid DDS.

OVT. This operation receives the list of face descriptor records produced by OVS and the geodetic data transmittal records passed from the image overlay process. It simply sorts them to produce a file containing a set of records for each new face in sequence of new face numbers.

OVU. This operation compiles and link-edits the source program generated in OVS above, producing a temporary load module for use in OVV. OVU refers to COVLIB and SYMLIB to obtain the DDS input-output areas and the OVV main line logic, respectively.

OVV. The purposes of this operation are to combine the descriptor data selected from the input DDS files with the geodetic data produced by the image overlay process, and to form a DDS record for each face in the new overlaid DDS. The data it needs for this process have been arranged on one file and passed to it from OVS. The program needed to carry out the process has been tailor-made by OVS and OVU to fit the types of data

involved and the size of the task. The records are processed sequentially and checked for error conditions, and a new DDS record is written for each new face.

P. Kingston and M. Doyle were responsible for the design and programming of the overlay operation.

#### PLOT

PLOT is a subroutine that accepts IDS format records and prepares a tape for use on a flat-bed plotter.

#### EXAMPLES OF OUTPUT FROM THE CANADA GEOGRAPHIC INFORMATION SYSTEM

Five examples of CGIS output are provided on the following pages. They demonstrate the capability of the system to: measure and tabulate data from maps of the same data type; compare maps of different data types and list areas with specified multiple characteristics; draw a new map showing selected land types in proximity to other features; perform multiple comparisons between maps of different data types and draw a new map showing only those combinations of land of interest to the user.

In each case the request being made is written as it would be initially stated by the user. Following the request are the samples of the output from the CGIS system made in response to that request. In the last example, copies of the original manuscript maps are provided so that the system output can be compared with the source documents.

These examples were generated by the Department of the Environment, Government of Canada, in 1972.

Example 1 - Area totals from maps of single data type.

Request: List all the unique present land uses classified in the data bank. Give the total extent of each type in acres and give the percentage of the total area that each type occupies.

*****LISTING OF UNIQUE PRESENT LAND USES*****			
CLASSES	DESCRIPTION	ACREAGE	PERCENT (LAND AREA)
A	95K-100K CROPLAND	0	0.0
A	75K-94.9K CROPLAND	0	0.0
A	50K-74.9K CROPLAND	17,869	1.4
P	95K-100K IMPROVED PASTURE AND FORAGE CROPS	493	0.0
P	75K-94.9K IMPROVED PASTURE AND FORAGE CROPS	15,434	1.2
P	50K-74.9K IMPROVED PASTURE AND FORAGE CROPS	542,606	44.3
B	URBAN LAND USE (NON-AGRICULTURAL)	6,946	0.5
E	MINES, QUARRIES, SAND AND GRAVEL PITS	702	0.0
G	ORCHARDS AND VINEYARDS	351	0.0
M	CROPLAND (FIELD CROPS-GRAIN, VEGETABLES, ETC)	838	0.0
K	ROUGH GRAZING AND RANGELAND	74,453	6.0
L	UNVEGETATED SURFACES (ROCK)	453	0.0
M	SWAMP, MARSH OR BOG	23,203	1.8
O	URBAN OUTDOOR RECREATION (PARKS, ARENAS)	874	0.0
S	UNPRODUCTIVE LAND (ISLAND)	5,513	0.4
T	PRODUCTIVE WOODLAND	463,457	37.9
U	NON-PRODUCTIVE WOODLAND (SMALL TREES, BUSHES)	69,367	5.6
Z	WATER (OCEANS, LAKES, PONDS, RIVERS)	2,081,575	0.0
TOTAL AREA (ACRES)		3,305,265	
TOTAL AREA EXCLUDING WATER (ACRES)		1,222,589	

Table A. 1. Canada Geographic Information System - Example 1.

Example 2 - Overlay and comparison of maps of two different data types.

**Request:** Overlay the present land use map on the map of the capability of the land for agriculture.

List the major agricultural limitations of the land, satisfying the following requirements:

- (1) The land must have an agriculture capability of 3 or less.
- (2) The land must presently be unproductive, that is, it must be either non-productive woodland, unimproved pasture and range land, cropland, horticulture, or improved pasture and forage crops.
- (3) Only land areas of greater than 10 acres are to be considered.

<u>FACE</u>	<u>PRESENT LAND USE CLASSIFICATIONS</u>	<u>AREA</u>	<u>CENTROID(DEC.)</u>	<u>AGRI. CLASSIFICATIONS</u>
<u>ID CODE</u>	<u>DESCRIPTION</u>	<u>LACRES</u>	<u>LAT. LONG.</u>	<u>CODE CLASS LIMITATION</u>
1	U NON-PRODUCTIVE WOODLAND	33.67	48.83 64.61 3	T TOPOGRAPHY
2	K UNIMPROVED PASTURE AND RANGE LAND	11.37	48.80 64.60 3	T TOPOGRAPHY
3	U NON-PRODUCTIVE WOODLAND	16.48	48.80 64.60 3	T TOPOGRAPHY
4	K UNIMPROVED PASTURE AND RANGE LAND	17.20	48.80 64.60 3	T TOPOGRAPHY
5	U NON-PRODUCTIVE WOODLAND	17.68	48.82 64.60 3	T TOPOGRAPHY
6	X UNIMPROVED PASTURE AND RANGE LAND	11.02	48.82 64.60 3	T TOPOGRAPHY
7	U NON-PRODUCTIVE WOODLAND	13.11	48.81 64.54 3	T TOPOGRAPHY
8	P IMPROVED PASTURE AND FORAGE CROPS	14.49	48.83 64.61 3	F FERTILITY
9	P IMPROVED PASTURE AND FORAGE CROPS	10.16	48.84 64.62 3	F FERTILITY
10	U NON-PRODUCTIVE WOODLAND	25.86	48.81 64.58 3	F FERTILITY
11	P IMPROVED PASTURE AND FORAGE CROPS	14.85	48.82 64.58 3	F FERTILITY
12	U NON-PRODUCTIVE WOODLAND	10.34	48.82 64.58 3	F FERTILITY
13	P IMPROVED PASTURE AND FORAGE CROPS	41.67	48.82 64.58 3	F FERTILITY
14	K UNIMPROVED PASTURE AND RANGE LAND	30.30	48.82 64.59 3	F FERTILITY
15	P IMPROVED PASTURE AND FORAGE CROPS	11.48	48.82 64.59 3	F FERTILITY
16	K UNIMPROVED PASTURE AND RANGE LAND	18.70	48.83 64.59 3	F FERTILITY
17	P IMPROVED PASTURE AND FORAGE CROPS	16.47	48.83 64.60 3	F FERTILITY
18	P IMPROVED PASTURE AND FORAGE CROPS	27.53	48.84 64.61 3	F FERTILITY
19	P IMPROVED PASTURE AND FORAGE CROPS	62.89	48.83 64.60 3	F FERTILITY
20	K UNIMPROVED PASTURE AND RANGE LAND	25.91	48.83 64.59 3	F FERTILITY
21	P IMPROVED PASTURE AND FORAGE CROPS	28.37	48.83 64.59 3	F FERTILITY
22	P IMPROVED PASTURE AND FORAGE CROPS	17.37	48.84 64.59 3	F FERTILITY
23	K UNIMPROVED PASTURE AND RANGE LAND	10.70	48.84 64.59 3	F FERTILITY
24	P IMPROVED PASTURE AND FORAGE CROPS	15.27	48.85 64.60 3	F FERTILITY
25	K UNIMPROVED PASTURE AND RANGE LAND	23.98	48.86 64.60 3	F FERTILITY

Table A. 2a. Canada Geographic Information System - example 2.



FACE PRESENT LAND USE CLASSIFICATIONS AREA CENTROID COORDINATES AGRICULTURE CLASSIFICATIONS

NO. CODE	DESCRIPTION	(ACRES)	LAT.	LONG.	CODE	CLASS	LIMITATION
26	H HORTICULTURE	16.59	48.84	64.60	3	F	FERTILITY
27	K UNIMPROVED PASTURE AND RANGE LAND	15.14	48.85	64.60	3	F	FERTILITY
28	P IMPROVED PASTURE AND FORAGE CROPS	35.84	48.85	64.61	3	F	FERTILITY
29	P IMPROVED PASTURE AND FORAGE CROPS	19.60	48.83	64.56	3	F	FERTILITY
30	K UNIMPROVED PASTURE AND RANGE LAND	17.01	48.83	64.55	3	F	FERTILITY
31	K UNIMPROVED PASTURE AND RANGE LAND	10.10	48.83	64.56	3	F	FERTILITY
32	P IMPROVED PASTURE AND FORAGE CROPS	57.36	48.83	64.57	3	F	FERTILITY
33	K UNIMPROVED PASTURE AND RANGE LAND	15.81	48.83	64.57	3	F	FERTILITY
34	K UNIMPROVED PASTURE AND RANGE LAND	15.21	48.83	64.58	3	F	FERTILITY
35	U NON-PRODUCTIVE WOODLAND	37.27	48.83	64.56	3	F	FERTILITY
36	P IMPROVED PASTURE AND FORAGE CROPS	17.86	48.81	64.54	3	F	FERTILITY
37	P IMPROVED PASTURE AND FORAGE CROPS	42.28	48.81	64.53	3	F	FERTILITY
38	P IMPROVED PASTURE AND FORAGE CROPS	49.67	48.83	64.54	3	F	FERTILITY
39	P IMPROVED PASTURE AND FORAGE CROPS	32.23	48.83	64.53	3	F	FERTILITY
40	K UNIMPROVED PASTURE AND RANGE LAND	57.78	48.83	64.52	3	F	FERTILITY
41	U NON-PRODUCTIVE WOODLAND	11.42	48.83	64.54	3	F	FERTILITY
42	K UNIMPROVED PASTURE AND RANGE LAND	19.18	48.83	64.53	3	F	FERTILITY
43	K UNIMPROVED PASTURE AND RANGE LAND	15.87	48.84	64.59	3	F	FERTILITY
44	P IMPROVED PASTURE AND FORAGE CROPS	20.99	48.82	64.57	3	F	FERTILITY
45	U NON-PRODUCTIVE WOODLAND	11.73	48.81	64.50	3	T	TOPOGRAPHY
46	K UNIMPROVED PASTURE AND RANGE LAND	49.87	48.93	64.67	3	F	FERTILITY
47	K UNIMPROVED PASTURE AND RANGE LAND	10.62	48.91	64.61	3	F	FERTILITY
48	P IMPROVED PASTURE AND FORAGE CROPS	30.97	48.91	64.62	3	F	FERTILITY
49	K UNIMPROVED PASTURE AND RANGE LAND	10.08	48.91	64.63	3	F	FERTILITY
50	K UNIMPROVED PASTURE AND RANGE LAND	15.78	48.92	64.63	3	F	FERTILITY

Table A. 2b. Canada Geographic Information System - example 2 continued.

FACE PRESENT LAND USE CLASSIFICATIONS AREA CENTROID (DEG.) 66BI CLASSIFICATIONS

NO CODE	DESCRIPTION	(ACRES)	LAT	LONG	CODE	CLASS	LIMITATION
51 U	NON-PRODUCTIVE WOODLAND	43.27	48.92	64.62	3	F	FERTILITY
52 P	IMPROVED PASTURE AND FORAGE CROPS	16.90	48.92	64.62	3	F	FERTILITY
53 P	IMPROVED PASTURE AND FORAGE CROPS	28.68	48.92	64.63	3	F	FERTILITY
54 K	UNIMPROVED PASTURE AND RANGE LAND	10.02	48.94	64.64	3	F	FERTILITY
55 K	UNIMPROVED PASTURE AND RANGE LAND	14.65	48.89	64.60	3	F	FERTILITY
56 P	IMPROVED PASTURE AND FORAGE CROPS	20.77	48.90	64.63	3	F	FERTILITY
57 K	UNIMPROVED PASTURE AND RANGE LAND	15.97	48.92	64.61	3	F	FERTILITY
58 U	NON-PRODUCTIVE WOODLAND	15.43	48.93	64.58	3	F	FERTILITY
59 P	IMPROVED PASTURE AND FORAGE CROPS	54.63	48.90	64.61	3	F	FERTILITY
60 A	CROPLAND	15.66	48.91	64.61	3	F	FERTILITY
61 K	UNIMPROVED PASTURE AND RANGE LAND	11.35	48.89	64.56	3	F	FERTILITY
62 K	UNIMPROVED PASTURE AND RANGE LAND	64.02	48.89	64.55	3	F	FERTILITY
63 U	NON-PRODUCTIVE WOODLAND	21.73	48.89	64.56	3	F	FERTILITY
64 K	UNIMPROVED PASTURE AND RANGE LAND	21.37	48.92	64.57	3	F	FERTILITY
65 K	UNIMPROVED PASTURE AND RANGE LAND	114.39	48.90	64.58	3	F	FERTILITY
66 P	IMPROVED PASTURE AND FORAGE CROPS	21.68	48.89	64.53	3	F	FERTILITY
67 P	IMPROVED PASTURE AND FORAGE CROPS	26.19	48.88	64.53	3	F	FERTILITY
68 U	NON-PRODUCTIVE WOODLAND	11.35	48.89	64.52	3	F	FERTILITY
69 K	UNIMPROVED PASTURE AND RANGE LAND	12.07	48.89	64.52	3	F	FERTILITY
70 K	UNIMPROVED PASTURE AND RANGE LAND	43.37	48.89	64.51	3	F	FERTILITY
71 U	NON-PRODUCTIVE WOODLAND	24.03	48.89	64.51	3	F	FERTILITY
72 U	NON-PRODUCTIVE WOODLAND	29.01	48.89	64.51	3	F	FERTILITY
73 U	NON-PRODUCTIVE WOODLAND	14.89	48.89	64.51	3	F	FERTILITY
74 U	NON-PRODUCTIVE WOODLAND	31.89	48.89	64.52	3	F	FERTILITY
75 P	IMPROVED PASTURE AND FORAGE CROPS	34.66	48.88	64.52	3	F	FERTILITY

Table A. 2c. Canada Geographic Information System - example 2 continued.



Example 3 - Identification of land characteristics adjacent to specific feature. Production of graphic output.

**Request:** Overlay the map of present land use on the map of land recreational capability in the vicinity of Lac de l'Est.

List the present land uses, the lands recreation capability, and the prime recreation feature of the area surrounding Lac de l'Est.

Plot and label all faces which are larger than 80 acres.

FACE PRESENT LAND USE CLASSIFICATIONS		AREA CENTROID(DEC.)		REG. CLASSIFICATIONS	
NO. CODE	DESCRIPTION	LACBES	LAT. LONG. CODE	CLASS	FEATURE
1 Z	WATER	.06	47.21 69.57	2	B BEACH
2 B	UNMAPPED AREA	.06	47.18 69.51	6	Q TOPOGR. PATTERNS
3 B	UNMAPPED AREA	.06	47.18 69.51	6	Q TOPOGR. PATTERNS
4 B	UNMAPPED AREA	.06	47.17 69.52	6	Q TOPOGR. PATTERNS
5 B	UNMAPPED AREA	.06	47.16 69.53	3	N LODGING
6 B	UNMAPPED AREA	.06	47.16 69.53	2	B BEACH
7 B	UNMAPPED AREA	.06	47.16 69.53	2	B BEACH
8 B	UNMAPPED AREA	.06	47.15 69.54	6	Q TOPOGR. PATTERNS
9 B	UNMAPPED AREA	.06	47.14 69.55	6	Q TOPOGR. PATTERNS
10 B	UNMAPPED AREA	.06	47.14 69.56	6	Q TOPOGR. PATTERNS
11 B	UNMAPPED AREA	.18	47.16 69.53	2	B BEACH
12 B	UNMAPPED AREA	.26	47.17 69.52	6	Q TOPOGR. PATTERNS
13 T	PRODUCTIVE WOODLAND	.55	47.19 69.50	8	UNCLASSIFIED
14 T	PRODUCTIVE WOODLAND	.93	47.22 69.57	3	B BEACH
15 B	UNMAPPED AREA	1.05	47.17 69.52	6	Q TOPOGR. PATTERNS
16 B	UNMAPPED AREA	1.55	47.13 69.56	6	Q TOPOGR. PATTERNS
17 B	UNMAPPED AREA	1.73	47.13 69.56	6	Q TOPOGR. PATTERNS
18 B	UNMAPPED AREA	1.66	47.18 69.52	6	Q TOPOGR. PATTERNS
19 B	UNMAPPED AREA	2.85	47.17 69.52	3	N LODGING
20 B	UNMAPPED AREA	3.47	47.19 69.51	6	Q TOPOGR. PATTERNS
21 B	UNMAPPED AREA	4.53	47.14 69.55	6	Q TOPOGR. PATTERNS
22 T	PRODUCTIVE WOODLAND	5.33	47.23 69.58	2	B BEACH
23 T	PRODUCTIVE WOODLAND	5.64	47.23 69.58	2	B BEACH
24 P	IMPROVED PASTURE AND FORAGE CROPS	5.92	47.22 69.57	2	B BEACH
25 B	UNMAPPED AREA	5.96	47.15 69.54	6	Q TOPOGR. PATTERNS
26 P	IMPROVED PASTURE AND FORAGE CROPS	10.35	47.21 69.57	3	B BEACH

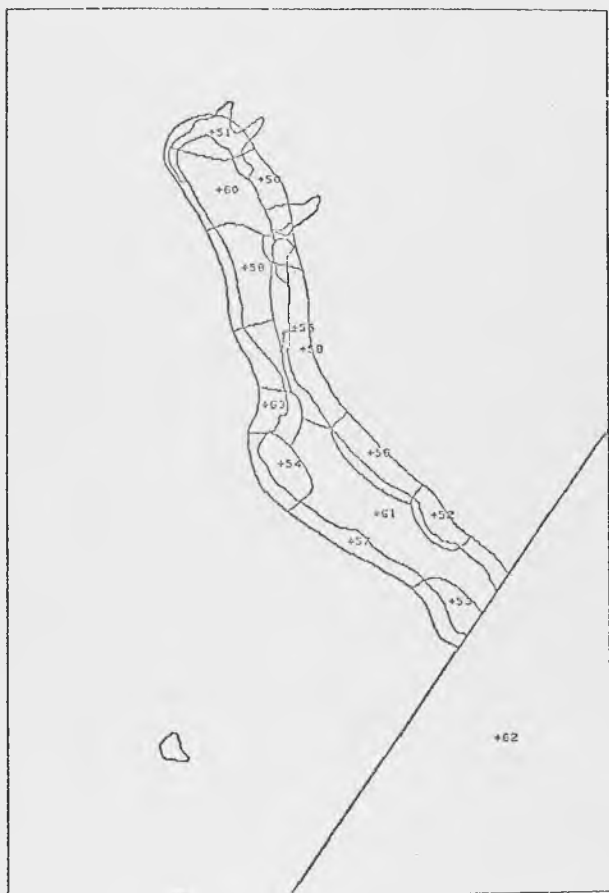
Table A. 3a. Canada Geographic Information System - example 3.

FACE PRESENT LAND USE CLASSIFICATIONS AREA CENTROID(REQ.) REG. CLASSIFICATIONS

LN CODE	DESCRIPTION	(ACRES)	LAT.	LONG.	CODE	CLASS	FEATURE
27	B URBAN: BUILT-UP AREA	10.53	47.24	69.58	6	Q	TOPOGR.PATTERNS
28	B URBAN: BUILT-UP AREA	12.77	47.23	69.57	6	Q	TOPOGR.PATTERNS
29	Z WATER	13.76	47.22	69.57	2	B	BEACH
30	B URBAN: BUILT-UP AREA	14.32	47.23	69.59	3	B	BEACH
31	B URBAN: BUILT-UP AREA	15.93	47.23	69.58	3	B	BEACH
32	T PRODUCTIVE WOODLAND	21.21	47.22	69.57	2	B	BEACH
33	T PRODUCTIVE WOODLAND	24.61	47.23	69.59	2	B	BEACH
34	P IMPROVED PASTURE AND FORAGE CROPS	24.80	47.22	69.57	2	B	BEACH
35	P IMPROVED PASTURE AND FORAGE CROPS	29.64	47.22	69.57	3	B	BEACH
36	Z WATER	30.10	47.18	69.54	2	B	BEACH
37	M SWAMP, MARSH OR BCG	30.24	47.15	69.59	6	Q	TOPOGR.PATTERNS
38	P IMPROVED PASTURE AND FORAGE CROPS	30.57	47.22	69.56	6	Q	TOPOGR.PATTERNS
39	T PRODUCTIVE WOODLAND	35.96	47.22	69.59	3	B	BEACH
40	Z WATER	45.26	47.23	69.58	2	B	BEACH
41	Z WATER	52.25	47.19	69.57	4	D	DEEPISSHORE WATER
42	T PRODUCTIVE WOODLAND	56.04	47.17	69.53	3	N	LOGGING
43	Z WATER	57.02	47.18	69.55	4	B	BEACH
44	Z WATER	67.02	47.20	69.57	2	B	BEACH
45	T PRODUCTIVE WOODLAND	71.61	47.18	69.57	3	B	BEACH
46	T PRODUCTIVE WOODLAND	71.97	47.19	69.57	4	D	DEEPISSHORE WATER
47	T PRODUCTIVE WOODLAND	72.32	47.16	69.54	2	B	BEACH
48	T PRODUCTIVE WOODLAND	74.42	47.21	69.58	4	D	DEEPISSHORE WATER
49	T PRODUCTIVE WOODLAND	75.43	47.20	69.57	2	B	BEACH
50	T PRODUCTIVE WOODLAND	80.42	47.23	69.57	3	B	BEACH
51	B URBAN: BUILT-UP AREA	86.30	47.23	69.58	2	B	BEACH
52	T PRODUCTIVE WOODLAND	87.01	47.18	69.54	2	B	BEACH

Table A. 3b. Canada Geographic Information System - example 3 continued.





DATE: FEB. 8/72  
JOB: D321092R

Fig. A. 54. Example 3 - CGIS graphic output.



Example 4 - Overlay and comparison of maps of two different data types.

Request: Overlay the map of present land use on the map showing the capability of the land for forestry.

How much land that has a forestry potential of Class 4 to 7 is currently being used as:

K - Rough grazing and rangeland

T - Productive woodland

U - Non-productive woodland (small trees, bushes)

M - Swamp, marsh or bog

S - Unproductive land (sand)

L - Unvegetated surfaces (rock)

ACREAGE IN PRESENT LAND USE CATEGORIES T.V.M. (FORESTED LAND)

FORESTRY  
CLASS  
CODES

	2	3	4	5	6	7	TOTALS
4	10,589	15,157	3,125	4,982	0	1,827	35,680
5	2,534	3,099	2,604	3,121	0	1,915	13,272
6	35	21	0	33	0	0	89
7	304	50	68	362	0	655	1,442
TOTALS	13,464	18,327	5,797	8,498	0	4,397	50,483

\*\*\*NOTE-TOTALS MAY VARY BY PLUS OR MINUS 3 ACRES BY ROWS OR 2 ACRES ON COLUMNS

\*\*\*\*\* ALL VALUES ARE IN ACRES

PERCENTAGE DISTRIBUTION ON COLUMNS

	2	3	4	5	6	7
4	79	83	54	59	0	42
5	19	17	45	37	0	44
6	0	0	0	0	0	0
7	2	0	1	4	0	15

PERCENTAGE DISTRIBUTION ON ROWS

	2	3	4	5	6	7
4	30	42	9	14	0	5
5	19	23	20	24	0	14
6	0	24	0	37	0	0
7	21	3	5	23	0	45

Table A.4. Canada Geographic Information System - example 4.

Example 5 - Multiple overlays. Weighted comparison. List and plotted output of selected areas.

Request: Select the highest weighted capability of an area based on the set of priorities shown in Table A.5. In the analysis, consider maps of the following data types:

- (1) recreation
- (2) agriculture
- (3) forestry
- (4) wildlife - ungulates
- (5) wildlife - waterfowl

A composite "coverage" is formed from an overlay of maps of the above data types.

The descriptor data for each area (face) in the composite coverage is scanned to determine which coverage has the highest capability. The descriptor data is then changed to reflect the classification of the highest capability only.

The resultant descriptor data is rescanned and if adjacent faces have the same descriptor, their common boundaries are removed and the faces are merged.

Figures A.55, A.56, A.57, A.58, and A.59 show the original input documents used as input to the first overlay. Although the entire map sheet was processed only the bottom right hand quarter is shown.

Output: The figure labelled "Selected High Potential Capability Areas" (Fig. A.60) shows the high capability subset of the coverage resulting from the above manipulations.

The listing following (Table A.6) describes the high capability areas for the entire map sheet. The label number on the plot is described by the corresponding face number in the listing.

CAPABILITY	PRIORITY	H.W.C.* CODE	AGRICULTURE AG	FORESTRY FO	RECREATION RE	UNGULATES UN	WATERFOWL WA	DERIVED CLASSIFICATIONS DESCRIPTION
HIGH	1	RE1			1			
	2	UN1W				1W		
	3	WA1					1	
	4	UN1				1		
	5	UN2W				2W		
	6	RE2			2			
	7	WA2					2	
	8	UN2				2		
	9	FO3		3				
MODERATE	10	UF3W						UNGULATES AND FORESTRY
	11	UR3W						UNGULATES AND RECREATION
	12	UN3W				3W		
	13	RE3			3			
	14	WA3M					3M	
	15	WA3					3	
	16	UF3						UNGULATES AND FORESTRY
	17	UR3						UNGULATES AND RECREATION
	18	UN3				3		
	19	FO4		4				
	20	RE4			4			
	21	AG4	4					
	22	WA4					4	
	23	NR4						AGRICULTURE AND UNGULATES
	24	UN4				4		
	25	FO5		5				
	26	AG5	5					
	27	RE5			5			
	28	WA5					5	
	29	NR5						AGRICULTURE AND UNGULATES
	30	UN5				5		
LTD.	31	UCUM						UNCLASSIFIED AND UNMAAPPED
	32	UCUR						UNCLASSIFIED (URBAN BUILT-UP AREA)
	33	UCNT						WATER
	34	ND6						NO DOMINANT USE
	35	NR7						ALL OTHER POSSIBILITIES (IF ANY)

\* HIGHEST WEIGHTED CAPABILITY

Table A. 5. Weighting factors - examples

The nomenclature used in this table is described on page 423.

# Nomenclature Used in Table A.5

Each data type and capability descriptor will be shown in the form "aabb", where "aa" specifies data type, according to the code:

Agriculture	AG
Forestry	FO
Recreation	RE
Ungulates	UN
Waterfowl	WA
Native range	NR
No dominant use	ND
Ungulates and moderate forestry	UF
Ungulates and moderate recreation	UR

and "bb" specifies capability within the data type, for example, 1, 1S, 1W, etc.

Native range coordinate. This coordinate may have two values, defined:

If Ungulates 4 and Agriculture 6 or 7, then -

Native Range capability = NR4

NR4 = ((UN4) and (AG6)) or ((UN4) and (AG7))

If Ungulates 5 and Agriculture 6 or 7, then -

Native Range capability = NR5

NR5 = ((UN5) and (AG6)) or ((UN5) and (AG7))

No dominant use. Where no single best capability in a given unit exceeds class 6, that unit shall then be classed "no dominant use".

## Ungulates and forestry.

If Ungulates (3W) and Forestry (4), then -

Ungulates and Forestry capability = UF3W

UF3W = UN3W and FO4

If Ungulates (3) and Forestry (4), then -

Ungulates and Forestry capability = UF3

UF3 = UN3 and FO4

## Ungulates and recreation.

If Ungulates (3W) and Recreation (4), then -

Ungulates and Recreation capability = UR3W

UR3W = UN3W and RE4

If Ungulates (3) and Recreation (4), then -

Ungulates and Recreation capability = UR3

UR3 = UN3 and RE4

Fig. A. 55.  
RECREATION  
605

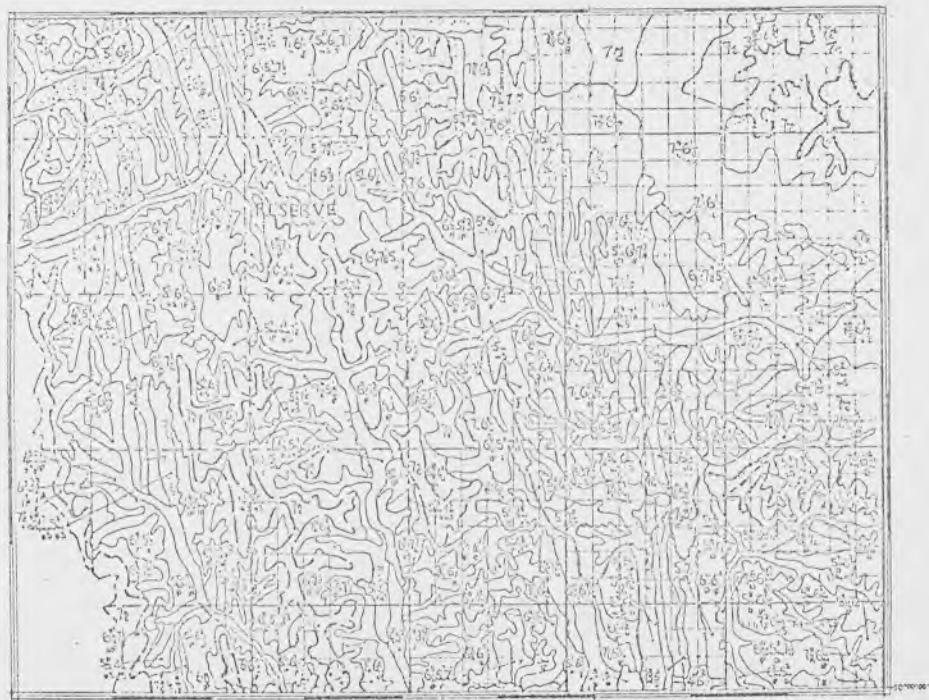
821 114-10-100

[illegible]

Fig. A. 56

AGRICULTURE

Original  
manuscript  
copy



KANANASKIS  
LAKES  
British Columbia  
Alberta

Scale  
1: 250,000

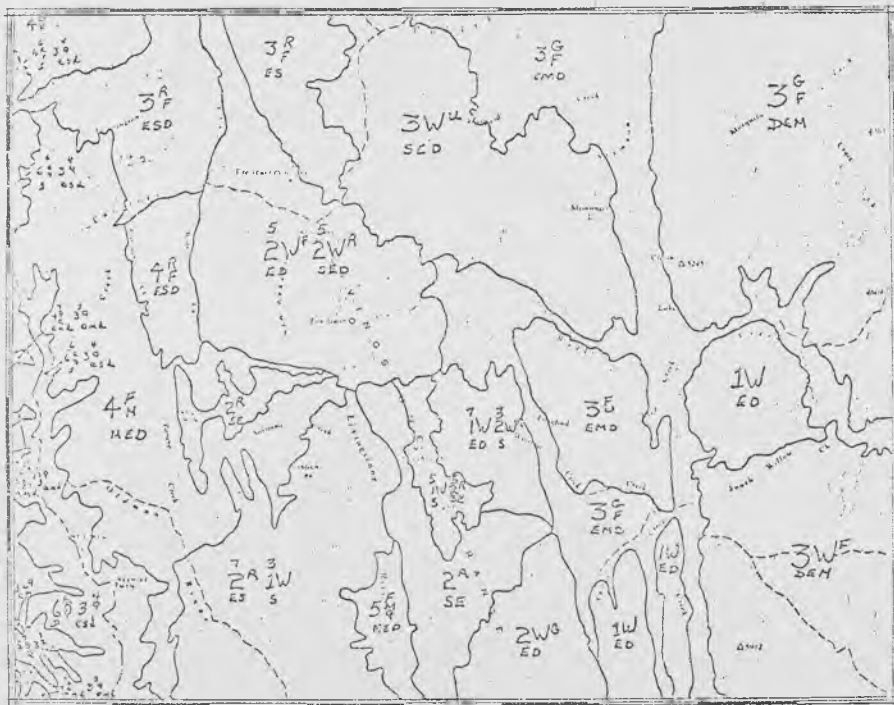
Fig. A. 57.  
FORESTRY  
305  
Original  
manuscript  
copy



KANANASKIS  
LAKES  
British Columbia  
Alberta

Scale  
1 : 250,000

Fig. A.58.  
WILDLIFE  
UNGULATES  
405  
Original  
manuscript  
copy



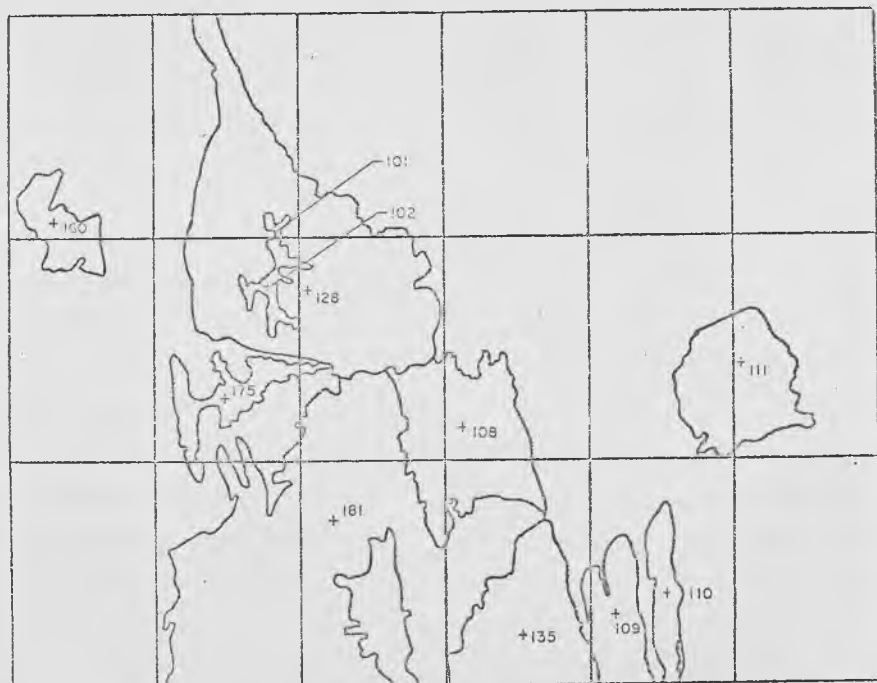


KANANASKIS  
LAKES  
British Columbia  
Alberta

Scale  
1 : 250,000

Fig. A.59  
WILDLIFE  
WATERFOWL  
505  
Original  
manuscript  
copy

821 114/90 00



SELECTED HIGH POTENTIAL CAPABILITY AREAS

82J

Fig. A.60. Example 5 - CGIS graphic output.



HIGH CAPABILITY				
FACE NUMBER	GEODETIC AREA (ACRES)	CENTROID		CLASSIFICATION DATA
		LATITUDE	LONGITUDE	
100	4,599.52	50.256	114.765	RE2
159	3,542.41	50.841	115.141	RE2
146	769.18	50.939	115.321	RE2
155	493.26	50.982	115.362	RE2
149	344.49	50.811	115.174	RE2
156	248.57	50.878	115.362	RE2
145	223.53	50.939	115.335	RE2
144	163.99	50.894	115.263	RE2
154	162.74	50.887	115.384	RE2
147	132.84	50.613	115.670	RE2
150	109.54	50.834	115.216	RE2
148	56.11	50.675	115.077	RE2
153	14.25	50.816	115.215	RE2
152	2.65	50.829	115.225	RE2
151	2.48	50.843	115.239	RE2
TOTAL ACPPAGE FOR THE ABOVE SINGLE BEST CAPABILITY CLASSIFICATION IS 19,806.28				

Table A. 6b. Canada Geographic Information System - example 5 continued.

HIGH CAPABILITY				
FACE NUMBER	GEODETIC AREA (ACRES)	CENTROID		CLASSIFICATION
		LATITUDE	LONGITUDE	DATA
106	19,543.42	50.743	114.776	UNIW
107	18,149.87	50.844	114.622	UNIW
108	16,833.24	50.	114.358	UNIW
111	12,563.21	50.	114.121	UNIW
109	10,589.08	50.	114.631	UNIW
112	9,705.67	50.	114.345	UNIW
109	6,580.67	50.	114.227	UNIW
110	4,040.54	50.	114.186	UNIW
103	3,856.18	50.	115.198	UNIW
TOTAL ACREAGE FOR THE ABOVE SINGLE BEST CAPABILITY CLASSIFICATION IS 101,859.68				

Table A. 6c. Canada Geographic Information System - example 5 continued.

HIGH CAPABILITY				
FACE NUMBER	GENETIC AREA (ACRES)	CENTROID LATITUDE	CENTROID LONGITUDE	CLASSIFICATION DATA
174	98,974.04	50.594	114.863	UN2
181	60,814.37	50.071	114.473	UN2
180	16,332.69	50.674	114.748	UN2
172	6,983.42	50.935	114.978	UN2
175	4,978.04	50.160	114.563	UN2
166	4,777.42	50.838	114.923	UN2
167	3,165.93	50.907	115.239	UN2
165	2,980.76	50.940	115.155	UN2
164	2,779.60	50.956	115.213	UN2
168	2,451.11	50.726	115.198	UN2
179	2,193.14	50.888	114.807	UN2
163	1,517.04	50.648	115.163	UN2
162	1,602.75	50.612	115.202	UN2
TOTAL ACREAGE FOR THE ABOVE SINGLE BEST CAPABILITY CLASSIFICATION IS 208,950.31				

Table A. 6d. Canada Geographic Information System - example 5 continued.

FACE NUMBER	GEODETIC AREA (ACRES)	CENTROID		CLASSIFICATION DATA
		LATITUDE	LONGITUDE	
128	41,629.96	50.243	114.502	UN2W
139	38,607.07	50.684	114.359	UN2W
124	21,540.75	50.845	115.069	UN2W
129	16,144.02	50.648	114.847	UN2W
135	14,147.35	50.033	114.306	UN2W
132	9,300.65	50.863	114.867	UN2W
126	8,141.95	50.455	114.748	UN2W
141	7,980.07	50.913	114.276	UN2W
140	5,139.62	50.513	114.124	UN2W
118	5,134.43	50.449	114.291	UN2W
131	4,102.14	50.760	114.876	UN2W
119	3,902.56	50.787	114.323	UN2W
136	3,648.12	50.565	114.291	UN2W
121	2,197.62	50.938	115.210	UN2W
143	1,057.25	50.868	114.712	UN2W
116	1,726.80	50.662	114.964	UN2W
142	883.76	50.849	114.414	UN2W
120	350.93	50.630	114.359	UN2W
114	22.07	50.469	114.720	UN2W
115	1.33	50.390	114.576	UN2W
117	.92	50.612	114.814	UN2W
113	.46	50.414	114.732	UN2W
TOTAL ACREAGE FOR THE ABOVE SINGLE BEST CAPABILITY CLASSIFICATION IS 185,149.83				

Table A. 6e. Canada Geographic Information System - example 5 continued.



HIGH CAPABILITY				
FACE NUMBER	GEODETTIC AREA (ACRES)	CENTROID LATITUDE	CENTROID LONGITUDE	CLASSIFICATION DATA
161	527.56	50.882	114.655	WA2
TOTAL ACREAGE FOR THE ABOVE SINGLE BEST CAPABILITY CLASSIFICATION IS				527.56

Table A. 6f. Canada Geographic Information System - example 5 continued.

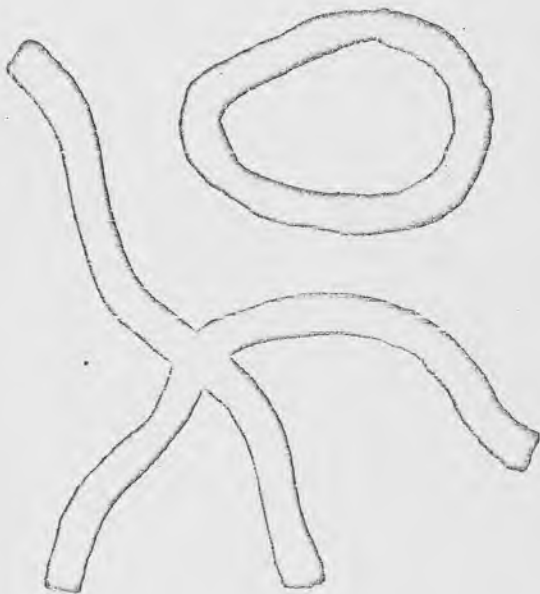


SUMMARY OF THE HIGHEST WEIGHTED CAPABILITIES	
HIGHEST WEIGHTED CAPABILITY	ACREAGE
WA2	527.56
UN2H	185,149.83
UN2	208,950.31
UN1W	101,855.88
RE2	10,006.28
RE1	2,199.54
FO3	778.88
THE TOTAL AREA PROCESSED IS	510,268.68 ACRES

Table A. 6h. Canada Geographic Information System - example 5 continued.

EXAMPLES OF V-VALUE ALLOCATION, CLOUD ELIMINATION,  
AND BOUNDARY TRACING

Source: Morton, G. M. 1966. Computer mapping - A map reduction technique. Unpublished working paper.



This is a small piece of the scribed image of the source map, as seen by the scanner. The view is highly magnified. In reality, the lines are only 0.008 in. wide.

Fig. A. 61. Input map.



This is the pattern of 0 and 1 bits produced by the scanner as a representation of the image that it scanned. (The 0 bits have been left blank in this diagram.) This 0 and 1 pattern is stored on magnetic tape. The graphic to non-graphic conversion has thus been carried out.

Fig. A.62. Scanner output.

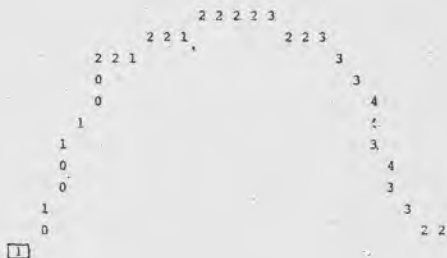


This is the result of performing the "V-value" calculation upon each of the bits of data describing the image now stored in the computer. Note that the higher the value now assigned to each bit, the closer it is to the center of the line.

Fig. A. 63. Result of V-value calculations.

DIRECTION CODES:

7	0	1
6	*	2
5	4	3

START POINT 1

Cloud elimination is the selection of points near the center of lines and the elimination of other points. The process is begun by choosing a start point with a suitably high V-value. Then the path of high V-values is followed. As the line is followed, the path is "marked" so that it will not be retraced. Simultaneously, a new map image is created elsewhere in core with only the selected points. The diagram above shows this new map. Note that the values are no longer V-values; they are actually direction codes which indicate the direction of the next point in the line. The line is traced in this manner until the edge of a section is reached, a previously traced line is encountered, or a line terminates itself. In the case above, the line was simply traced until the edge of the section was found.

Fig. A. 64. Result of cloud elimination - first line.



The tracing process scans the section until another suitably high V-value start point is found. The process is repeated until all paths in the section have been traced. Intersections are recognized and marked when previously traced lines in the section are met:  $\Delta$ . Given the line image now identified, it is a simple matter to line-follow the direction codes along each line segment and convert each point on the line to an X-Y coordinate, and record the coordinates in a table. During this latter line-following process, "colours" (unique numbers) are assigned to each area and small gaps in the lines are connected.

Fig. A. 65. Result of cloud elimination - remaining lines.



PARTIAL LIST OF PERSONS INSTRUMENTAL IN THE DEVELOPMENT  
OF THE CANADA GEOGRAPHIC INFORMATION SYSTEM  
IN APPROXIMATE ORDER OF APPEARANCE

1961 R. F. Tomlinson  
W. G. E. Brown  
J. Sharp  
L. E. Pratt  
H. Lerchs  
A. Aus  
J. Thacker  
I. Grossman  
M. Ray  
J. Lewis  
C. Porter  
G. Morton  
J. Russell  
H. Knight  
D. Lever  
D. Thompson  
R. Kemeny  
A. Benjamin  
F. Jankaluk  
P. Kingston  
B. Sparks  
E. Beaudette  
C. Brown  
R. Predznovik  
W. Switzer  
W. Storey  
A. K. Millar  
J. Bolton  
B. Ferrier  
G. Fierheller  
T. Lakatos  
J. White  
M. Doyle  
K. Ward  
G. Leismer  
B. Giles  
R. Shaw  
J. Medd  
B. Cook  
D. Pearce  
P. Hatfield  
D. Lee  
R. Whittaker  
P. Bedard  
J. Foster

1969

Plus current Department of the Environment staff of  
approximately 10 persons.

## CANADA GEOGRAPHIC INFORMATION SYSTEM BIBLIOGRAPHY

Published Papers

- Kingston, P. P. C. H. 1968. A computer oriented geographic information processing system. Paper presented at the 10th Annual Conference of the Canadian Operational Research Society, Toronto, Ont. May, 1968.
- Thompson, D. R. 1967. A cartographic scanner. Proceedings of the 27th Annual Meeting of the American Society of Photogrammetry - American Congress on Surveying and Mapping.
- Tomlinson, R. F. 1962. Computer mapping: An introduction to the use of electronic computers in the storage, compilation and assessment of natural and economic data for the evaluation of marginal lands. Report presented to the National Land Capability Inventory Seminar, Agricultural Rehabilitation and Development Administration of the Canada Department of Agriculture.
- \_\_\_\_\_. 1963. Feasibility report of computer mapping system. Spartan Air Services Ltd., Ottawa, Ont.
- \_\_\_\_\_. 1967. An introduction to the geographic information system of the Canada Land Inventory. Department of Forestry and Rural Development, Ottawa, Ont.
- \_\_\_\_\_. 1968. A geographic information system for regional planning. In Stewart, G. A., ed. Land evaluation. Papers of a CSIRO Symposium organized in cooperation with UNESCO, Canberra, 26-31 August 1968, pp. 200-210. Macmillan of Australia, South Melbourne, Victoria.

Unpublished Working Papers and Notes

- Jankaluk, F. 1967. A reference coordinate system for computer mapping.\*
- \_\_\_\_\_. 1967. A system design for a geographic information system.\*
- Kemeny, R. L. 1968. A generalized method for representing digitized line data in compact form for computers.\*
- Kingston, P. P. C. H. 1968. Methods of overlaying digitizer line data in core.\*
- Morton, G. M. 1966. Computer mapping - A map reduction technique.\*
- Tomlinson, R. F. 1964. Canada Land Inventory - Economic feasibility report of data coordination methods.
- \_\_\_\_\_. 1969. Information from the data bank - Retrieval capabilities of the Canada Geographic Information System.+

Unpublished System Documentation

Program documentation, detailed flow charts, operators' and programmers' guides and manuals exist as working documents and are being continually revised. They are regarded as in-house working records and are not in a form suitable for publication or for use by persons other than those actively developing the CGIS system.

\* Written for IBM technical information exchange purposes.

+ Prepared for internal government information exchange purposes.

PUBLISHED CONTRIBUTION

Reprinted from Stewart, G. A., ed. 1968. Land evaluation.  
Macmillan of Australia.

## A Geographic Information System for Regional Planning

R. F. Tomlinson

*Department of Forestry and Rural Development,  
Government of Canada*

---

*As a tool in its program of rural development, Canada is developing a computer-based information system for the storage and manipulation of map-based land data. The system and its capabilities are described.*

Canada, like many countries, faces an immense problem in both understanding and guiding the development of its land, water, and human resources. One of the major agencies created specifically to implement policy to attack this problem is the Rural Development Branch of the Department of Forestry and Rural Development. A primary task facing this agency is to assemble social (demographic), economic, and land data for an integrated analysis to enable problems of rural development to be specified, development programs to be implemented, and their effectiveness evaluated.

Parallel with the gathering of data has been the development, by the Regional Planning Information Systems Division of the Branch, of interrelated computer-based information systems to handle and analyse the data. The Geographic Information System, for the storage and manipulation of land data is the most developed of these systems. Its design and development started in 1963, implementation began in 1965, and is now in its final stages; routine use is scheduled for September 1968. It is perhaps worthwhile to recount our progress with this system at this time.

Early in the life of the Branch (1962) a start was made with the gathering of some kinds of land data by the Canada Land Inventory. The data they collect is restricted to five types: the present use of the land, the capability of the land for agriculture, the capability of the land for forestry, the capability for recreation, and the capability for supporting wildlife. These data alone, if gathered in sufficient quantities for the summaries to be directly applicable to provincial and federal resource policy and regional planning, will generate an estimated 30,000 map sheets, at various scales. The Inventory has currently produced 7000 map sheets, of which 3000 have been prepared for computer input. The maps contain an average of 800 distinct areas on each sheet, and have been found to contain as many as 4000. Additionally, other types of maps covering watersheds, climate, geology, administrative boundaries, and land titles are generated by other agencies.

The need for a computer-based system, whereby map and related data can be stored in a form suitable for rapid measurement and comparison, is apparent as soon as the magnitude of the problem of handling large numbers of maps is appreciated. Lack of trained personnel makes it impossible to examine such large amounts of data manually in any sensible time, much less to provide a meaningful analysis of the content. A situation can be reached where the amount of data precludes its use. The end product of countless hours of survey can remain unused, with the result that administrators do not receive information necessary for a sound basis to decision making.

From the first, it was the intention to produce the maps generated for the Canada Land Inventory in such a way that their data could be related on a nation-wide basis by the geographic information system. This made it necessary to establish a common basis of data description. Classification systems were evolved for each type of data by discussions with the federal and provincial agencies concerned in the original survey, under the guidance of a federal coordinator. In each case, the classification systems were subject to trial in pilot areas in various regions of the country. Regional variations are incorporated into the classification system by development of ratings which recognize equivalent values. The classification systems vary from a relatively simple, one-letter code for present land use to a complex, multi-level description used for forestry.

The maps, essentially interpretations of existing data in terms of the classification system, are usually produced by the federal and provincial agencies most closely related to the collection of the original data (over 100 agencies are involved). The manuscript maps are sent to Ottawa to be edited and prepared for computer input.

The basic capability of the geographic information system is that it accepts and stores all types of location-specific information, that is, any information which can be related to an area, line, or point on a map. Information relating to land resources is most frequently location-specific in character. For example, census data (perhaps not usually thought of as location-specific) are collected from specific areas of land called enumeration areas, which are recorded on maps; a highway is a location-specific line; a campsite can be thought of as a location-specific point on a map.

The system can best be described as comprising two parts: the data bank and the set of procedures and methods for moving data into the bank, and for carrying out the manipulations, measurements, and comparisons of the data, once there. These two parts will be referred to as the 'data bank' and the 'information system', respectively. It is quite possible to have the entire geographic information system with full operating capability and have no data in the data bank. The amount of data which can be put into the data bank is infinite, as any number of magnetic tapes can be generated and stored. Additional data related to any area can be inserted at any time.

The system has the following capabilities:

It will accept maps containing data represented as areas or lines or points. The maps can be of any scale and on any map projection, and they can contain linear distortions. All of these characteristics will be adjusted to a standard format (normalized) when they are put in. Data relating to points only can be put in independently of maps. They are simply related to their latitude and longitude points.

The system compacts and stores information. The compaction is most efficient. For maps at a scale of 1:50,000 with an average density of information it is expected that a complete coverage of the farmed area of Canada (approximately 600 map sheets) can be recorded on two reels of magnetic tape.

The system can measure any data in the data bank. If the data have been inserted in the form of areas, then each area can be measured. For example, a soil map might be represented by different areas of different soils. The area of each patch of soil or the total area of any one type of soil can be calculated.

Similarly, the lengths of lines can be measured and the occurrences of points counted.

The region from which area, line, or point measurements are required can

be limited in a variety of ways. Data can be retrieved within any boundary already described to the system. If, for example, a map of administrative region boundaries has been put into the data bank, measurements can be carried out within a specific administrative region. If a desired boundary has not already been described to the system it can, of course, be drawn on a clean sheet and inserted in the normal way, or if it is simple enough in shape to be described by a straight line joining points, then it is only necessary to put in the co-ordinate values of the points.

It will also be possible to limit retrieval by reference to any line or point already described to this system. The system can be asked, for example, to measure the area of patches of land crossed by the line of a highway or within a band of specified width along the highway, or to determine the areas suitable for sub-divisions within 20 miles of the center of a city.

A major system capability is comparison of two types of mapped data relating to the same area. Just as two maps can be mutually overlaid to allow the relationships between the data to be examined, the system can overlay any two or more types of data to measure the exact amounts of each type of land in juxtaposition to the map or maps below.

This can be applied as a search capability, whereby a comparison of various types of information is made to find out where a selected set of characteristics occur together. For example, a request to find suitable landing sites for a helicopter would require an examination of the vegetation map to determine treeless areas, the topographic map to make sure that the area was flat, and the present land use map to make sure that the area was not populated. These three coverages would be compared to identify and describe all points having the desired characteristics.

A further extension of the search capability could result in a 'search in context'. A potential helicopter landing-site, for example, would be of limited value if, while being perfectly treeless, flat, and uninhabited, it occurred as an island in the middle of a swamp. The search routine can be instructed to ignore otherwise desirable sites if they do not occur in a desirable context.

Another search capability that can be implemented is referred to as the 'nearest neighbour search'. This would be employed when the limit of the search is not definite enough to be specified. The search command would simply request the nearest examples of the desired character to be located. A composite example of some of these capabilities might be an instruction to locate the nearest potash mine which is served by a main highway, north and south railroad connections, and is surrounded by a minimum of 10,000 sq miles of good farmland.

The system can produce information in two different forms. The commonest form is perhaps the normal printed alphabetical and numerical data produced on the regular computer printer. In addition to the printer will be a graphic plotter which, under the control of the system, produces a map showing the location of the desired areas, lines, or points which satisfy the request.

An inherent danger of information systems is that the data entered into the system may vary widely in reliability, but may be assumed to be equally reliable in subsequent multifactor assessments. The system can accept a reliability identifier with any type of information and can keep track of reliability tags so that degrees of reliability are printed out beside the answer to a request.

The advantages of information which is kept up to date, compared with data which have to accumulate for several years before it is economically desirable to reprint a map, are well known to users of map information. Data can easily be added to the system without waiting for large amounts of new data to

accrue. Old coverage can be erased and replaced on the magnetic tapes or, if desired, both the old and the new coverage can be retained. New survey data at a more detailed scale can be incorporated with previous data at smaller scales, provided, of course, that the classification systems are compatible.

For many of the day-to-day information needs of administrators of land resource policy, simple forms exist to allow the administrator to initiate the request without the assistance of a computer programmer. Although more detailed assessments requiring the full flexibility and capability of the system would best be handled by someone acquainted with the data formats, a considerable amount of programming effort has been eliminated even at this level by use of programs already written and incorporated into the system. It is estimated that, with no previous computer knowledge, an administrator could be taught to complete normal form-originated requests in one week. Three weeks training and practice thereafter are expected to be necessary for the same administrator to handle more detailed requests. The unusual or very complex requests will need a programmer working in conjunction with the system librarian.

In many ways the system is self-monitoring. On accepting a request for information, the first response of the librarian will be to use the system's KWIC<sup>1</sup> index to check whether that particular request has been made before and, if so, to indicate where the answer is stored in the filing cabinet. If the request has already been partially answered, this also is determined. If the request requires new manipulation of data, the system indicates which tapes have the requisite data stored on them.

The tapes then are selected from the library, put on to the computer and the assessment is executed. An extension of this capability is to provide a cost estimate of the work, prior to processing, based on a preliminary analysis of the amount of data on the requested tapes. Such estimates will be necessary in more complex applications.

The system is independent of peripheral devices such as input scanners or output plotters. While the IBM cartographic scanner is now in use, in conjunction with a D-Mac X-Y digitizer, to convert graphic data to digital form, instrumentation is likely to be developed in the next two or three years to combine these functions.

The normalization step, which converts digitized graphic information to the format required by the data bank, is independent of the main system functions and can be changed accordingly.

The system is designed for use on the IBM System 360 Model 50, with 512 thousand bytes<sup>2</sup> of storage, 6 magnetic tape drives, and 3 magnetic disc drives under the control of the standard operating system. Greater operating efficiency is achieved if the System 360 Model 65 is used. The practical application of the data bank concept and the entire system capability is available by use of this general-purpose computer.

#### SYSTEM DESCRIPTION

Boundary data to be put into the data bank are traced (scribed) on to a clean

<sup>1</sup> KWIC - Key Word In Context document indexing and cross-referencing system based on computer sorting of key words in the title. Ref. IBM Publ. E20-8091.

<sup>2</sup> Byte - a unit of computer storage space made up of eight digits, or bits, in the binary system (using only 0 or 1). Each byte is capable of storing one letter, two decimal digits, or a binary value.

sheet from the source map (Fig. 1). The unique areas or 'map elements' are numbered on a transparent overlay and the corresponding classification is transcribed to a data sheet for punching into cards to be read by the computer.

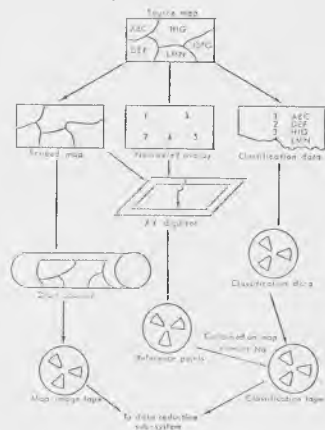


Fig. 1. Diagram showing flow of data preparation procedures.

The traced boundary sheet is placed on the drum scanner, and the scanning operation produces a digitized map of the boundaries on magnetic tape. The drum scanner was developed to meet Rural Development Branch requirements by the International Business Machines Company. The possible use of the drum scanning approach was first considered in 1963. The preliminary design criteria were established by the Rural Development Branch in 1964 and development work was contracted to the International Business Machines Company in 1965. The scanner consists of a cylindrical drum on which a map or chart can be mounted, and a movable carriage which slowly moves the scanning head across the front of the revolving drum. The scanning system consists of the scanning head proper, its associated electronics, and controls leading to a standard IBM magnetic tape drive.

The technique employed is to detect the changes in intensity of light reflected from black or white areas on the map or chart surface and to record this information as a series of binary bits written on magnetic tape. The scan head is a device utilizing fibre optics and is capable of scanning eight scan lines simultaneously. The drum scanner can accept a map up to 43 in. x 43 in. in size. A full-size map takes approximately 15 minutes to scan, including the time for mounting and dismounting it. Smaller sheets take a correspondingly shorter time.

It is not within the scope of this paper to give a detailed description of the drum scanner, though it is hoped that the engineering aspects will be covered in detail in a future paper. The format of the map-image data on tape is, however, pertinent to the discussion. One map-image record is produced for each 0.032 in. along the X-axis of a map sheet, and the height of each record area is 0.004 in. along the Y-axis. The 0.032 in. record, comprising one byte of computer storage, is divided into eight bits. Each bit thus represents an area or spot 0.004 in. wide. Lines drawn on the map are usually 0.008 in. wide. If the scan heads on the scanner identify 50%, or more of a spot as part of a line, then a '1' bit is generated; otherwise a '0' bit is generated. A line in this manner is represented as a collection of bits which are usually either one, two, or three spots in width.

The traced boundary sheet with the transparent numbered overlay is placed on a D-Mac cartographic X-Y digitizer where the four reference corner points and the co-ordinates of one reference point per 'map face' are coded in digits. A map face is any one of the distinct areas that together make up the surface of the map. As noted before, information related to a face is considered to be homogeneously distributed within that face. The output from the X-Y digitizer is produced on magnetic tape by means of an NCR encoder; this will revert to punched cards if it is found that the error-edit capability of cards is needed. The classification data sheet is now also directly transcribed on to magnetic tape, though this may be taken back to punched card output. Classification data and the digitized reference points are combined on the basis of map face number to result in a classification tape.

#### *Entering Data into the System*

The basic approach to feeding map data into the system is to reconstruct a line segment, or the part of a line that lies between adjacent vertices, from the point comprising the scanned map image. These segments are then combined with the classification information to produce map faces which are a basic unit of storage.

The following are some of the steps in this input procedure. As a preliminary, the identification of the scanner and classification tapes, coverage and map identification, and similar data are put into the procedure which controls the flow in the subsequent update<sup>1</sup> operation. The classification tape is edited for data consistency and is changed into system format during this stage (Fig. 2).

The map-image tape then enters the main map-data reduction procedure. Since a 30-in. by 30-in. map generates over 56 million bits, occupying over 7 million bytes of computer storage on an IBM System 360, the data reduction of the map image is performed sequentially on smaller units known as 'sections'. The use of a square (or nearly square) section results in considerably longer lines being available from the map for processing at one time than would be the case if a long, thin rectangle were used. A computer with 512 thousand bytes of core storage can handle a section in the order of 14 in. x 2½ in.

Each spot in the cloud of spots which make up the lines is assigned a 'V' value. This is a measure of the number of information-carrying spots surrounding it.

<sup>1</sup> Update - A computer procedure to combine new data being entered into the system with data already existing in the system. This may take the form of correcting, replacing, or deleting existing data, or inserting or adding new data.



[illegible]

The center points are coded to identify line intersections (or vertices) and the sense of direction of the line. Having thus located the points which comprise boundary lines, it is a simple task to record the X and Y co-ordinates of each point along a segment.

The system requires descriptive information to be related to map elements. One method of accomplishing this is to apply an identifying tag to both sides of the line. This tag also indicates in which direction the line was first followed, this being necessary if the sides of the line are to have a constant meaning.

The identifying tags are called 'system colors'. They are analogous to the colors in a political map. A sort-and-search of these colors enables the segments to be connected with each other, and hence faces to be assembled.

Using the reference points in latitude and longitude taken from the four corner points of the map, a transformation is carried out which locates the X-Y digitizer map-element reference points within the scanner. Map projections, which can vary from one map to another, are normalized. Calculations are made to correct for linear distortion and skewed orientation of the map on the scanner or digitizer. The transformed 'map-image data set' and the classification (or 'descriptor-data set') then are matched and compacted. During this match-and-compact operation, the map-image co-ordinates are recorded in terms of a standardized geodetic co-ordinate system. This allows a uniform base for storage and the subsequent measurement and overlay procedures.

The choice of a standard co-ordinate system was a major consideration. The eventual measurement needs (i.e. area, length, and centroid) required the chosen system to be locally cartesian. However, a co-ordinate system based on a projection can result in a system of regions, each with its own co-ordinate system. This problem is particularly pertinent when one considers an area as extensive as Canada.

Careful investigation indicated that a system comprised of the geodetic latitude and longitude had many advantages. The smallest division in the geodetic co-ordinate system used in the data bank is called a unit grid. It represents an angular displacement of  $1/2^{\circ}$  degrees. This was derived quite empirically. Using a 4-byte unit, 1 byte allows a span of 128 degrees which is sufficient to encompass Canada. The remaining 3 bytes represent the possible subdivision of any one degree.

The theoretical resolution of the system is determined by the actual distance on the ground covered by this unit grid, which at 45 degrees latitude is just over 1.4 in. in the latitudinal (or X) direction. This is considered adequate for the data being put into the system.

Scale within the system is in terms of the unit grid distance. Factors from  $2^0$  to  $2^{24}$  have been devised to provide coarser resolution.

To handle map information within the system, it is convenient to subdivide the co-ordinate system into regions called 'frames'. A frame has an equal angular displacement in the X and Y directions, hence is a square in the geodetic co-ordinate system.

A relatively simple calculation reveals that a map of average density (30 in. by 30 in., with 800 in. of boundary lines), will occupy 200,000 bytes of storage if no scale change or transformation is performed. With up to 30,000 maps envisaged as the primary content of the data bank, a compact notation for storage of co-ordinates was essential.

With a code based on direction change between co-ordinates and distance between co-ordinates, a sequence of simple codes can be used to describe co-ordinates. A sample line, requiring 864 bits for normal X-Y recording, occupies 76 bits in compact notation. If required, lines with regular patterns can be further compacted by storing the pattern with an indication of how many times the pattern is repeated.

In the match-and-compact phase, routines are carried out to calculate the area of each face, the centroid of face elements, and the length of line elements. In the same phase, an extensive error analysis is performed to ensure that the map is topologically correct. Errors found at this stage are documented by a series of error messages on the computer printer.

The match-and-compact operation produces two index files. The first of

there is a face file with classification and frame number which, when sorted, is used in updating the descriptor-data set. The second is a face file with segment identifiers which is used to update the image-data set. Incorporated in the second file is the basic compact notation of co-ordinate data by frame number. The routine for updating the image data set provides the geodetic properties (area, centroid, and length) as required to update the descriptor-data set. Both of these update routines can produce error listings as new data are matched with data already in the data bank. Again, error correcting is carried out as an update to the primary map-data reduction phase.

The best approach to take with regard to error correction will only be found by trial with a working system. Given a high percentage of errors requiring reference back to source documents or even to field survey, the relatively expensive method using cathode-ray tube displays would add little, if anything, to the efficiency of the error-correction procedure. On the other hand, given a high percentage of errors of a strictly cartographic nature and not requiring reference to source documents, the cathode-ray tube approach, by which images displayed on the tube can be corrected by 'drawing' on it with a beam of light, would have considerable merit. Both approaches will be investigated during the system trials.

#### *Data Bank Organization*

The data bank is divided into classification data contained within the descriptor-data set and boundary data contained within the image-data set. Three levels of file organization are envisaged. These are: (1) consecutive, (2) regional, and (3) indexed. These file organizations, together with an unstructured or structured version of the classification data within the descriptor-data set, have been combined into six levels. Five of these will be possible within the present scope of the data bank.

Using the descriptor-data set as an example, the relationship between the various levels can be thought of as follows: Level 1 represents the basic descriptor-data set arranged by consecutive face number; Level 2 represents a sorted Level 1, grouped according to some selected characteristic or set of characteristics; Level 3 is the equivalent to Level 1 for a specific region or group of regions; Level 4 can be thought of as a Level 3 which has been structured by grouping the faces relating to a certain characteristic or set of characteristics; Level 5 is a Level 2 or 4 which is not only structured but has an index of its contents available to facilitate further search. Level 6 is not implemented as an indexed consecutive file is not an advantage.

In the descriptor-data set for each map element, there is a list of pointers to the frames containing relevant parts of the boundary information for that map element. The format of this key varies with the level of file organization, but in all cases, it serves to relate the image-data set to the descriptor-data set. The record formats of the various levels of descriptor-data set are illustrated below.

#### *Level 1, 3*

Record Type	Coverage Number	Map Element	Geodetic Data	Factor Data	Frame List	Level 3 Region List
-------------	-----------------	-------------	---------------	-------------	------------	---------------------

*Level 2, 4, 6*

Record Type	DDS Key Classification Data	Geodetic Data	Non-Key Factor Data	Frame List	Region No. List
-------------	-----------------------------	---------------	---------------------	------------	-----------------

*Data Retrieval*

As the boundary information is kept separate from the description information it is only necessary to use the boundary information if actual boundaries have to be compared or output. Otherwise, all retrieval can be done from the description information files. This leads to extremely efficient use of the data bank, as most requests will not require use of the boundaries.

A computer needs a detailed description of the location and organization of data within itself before it can bring it out or manipulate it. These detailed descriptions are themselves kept in computer storage and are indexed by key words. These key words have been made to be the normal words that would ordinarily describe the maps such as PRESENT LAND USE OF AGRICULTURAL CAPABILITY. The use of such key words automatically generates computer programs that both describe the data and actually bring it out of the computer.

In the same way key words are used to describe the types of manipulation that can be carried out by the system. When data is to be retrieved the request is written by combining the key names of the data and the key words of the desired analysis. This results in a very powerful set of instructions being available that are also very flexible. This flexibility in data specification statements is made possible by use of the PL/I language. Uncomplicated requests will be extremely simple to address to the computer. The more complex requests will necessitate a small program being written, but even this will be facilitated by the use of these key words which represent already written small programs.

*Overlay Procedure*

The overlay procedure of the system is the well-known function of putting one map over another and examining the resulting data relationships.

Firstly, the two maps in the data bank are brought to the same scale. Then a section of one map of a size that can be handled by the computer is brought into core and the corresponding section of the map being overlaid is similarly brought into core and superimposed on the first. This, in effect, creates a new map with new faces. The new faces are 're-colored' and identified as new homogeneous areas. The first description data set is then brought in and the proper description is applied to each of the new faces. The description data set from the overlay map is similarly brought in and applied to the new map faces. Each of the new faces has now got a double name, one from each of the original two maps. The process is thus one of creating one 'new' map from the two original maps being overlaid. The new map can then have its areas measured and summarized in the same way as any other map in the system. It is stored and kept in the system as if it were an original map coverage. Up to eight maps can be overlaid in the same operation but obviously this is not a limitation, as the results of two overlays can subsequently be themselves overlaid.

*Data Control*

Data control within the system is achieved by the system monitor. The system monitor accepts pertinent data on the history of map-data manipulation within the system at all times. Many of the responsibilities for system control in such an open-ended system must rest with the system librarian.

The librarian's responsibilities include deciding whether coverages are permanent or temporary, selecting the resolution at which boundary lines for various coverages need to be stored, and deciding the way in which the descriptor-data sets are filed for ease of retrieval and comparison. He is also responsible for providing the procedures which edit the classification data in the preliminary phase of the map-data reduction sub-system. He must tailor the key words that describe the different types of map and different types of manipulation to efficient, specifically applicable retrieval requirements. He is in control of the flow of individual maps within the system and, similarly he must evaluate the practicability of assessment requests, including the avoidance of duplicate assessments.

## CONCLUSION

The Geographic Information System of the Rural Development Branch is still in an early stage of its development. Not all the procedures described have yet been fully implemented and at present rates of progress it will be several years before the data bank contains maps for any one type of information that cover the whole of the settled portion of Canada. The effectiveness of the system will of course depend as much on the quality of the data entered into the bank as on the capabilities for handling data. Nevertheless, the system is further advanced than any other major land data bank and contains several new concepts and techniques, especially those relating to the compact storage of boundary data and the rapid comparison of one map with another. Such a system is essential to effective rural planning in any country and offers for the first time the possibility of rapid and efficient geographical analysis which has application in any nation where the developing economy is concerned with the natural resources.

*The author gratefully acknowledges the support and encouragement given this work by ARDA, particularly from A. T. Davidson, J. E. Pratt, and W. A. Benson, Chief, Canada Land Inventory; and from the International Business Machines Company of the IBM, Ottawa staff associated with the programming and computer applications of the project; specifically: Guy Morton, Frank Jankulak, Don Lever, Art Benjamin, Robert Kenney, Peter Kingston, and Bruce Ferrier.*